The Dissertation Committee for Katherine Elizabeth Haning Certifies that this is the approved version of the following Dissertation:

Uncovering and Engineering Complex Phenotypes in Bacteria Using Regulatory Noncoding RNAs

Committee:

Lydia Contreras, Supervisor

Hal Alper

George Georgiou

R. Adron Harris

Uncovering and Engineering Complex Phenotypes in Bacteria Using Regulatory Noncoding RNAs

by

Katherine Elizabeth Haning

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2018

Acknowledgements

When I began this work, I knew nothing of bacteria or RNA or why I should care about them. Over time, I absorbed a great deal of understanding from my lab mates and through personal study. I started to catch on to the amazing intricacies of how living things operate.

I am especially thankful for my advisor Lydia Contreras's curiosity and passion in this field. It is contagious in the best way. She was so generous with time and resources to keep me going in the right directions. I learned a ton from Lydia about how to teach, mentor, foster a collaborative environment, and communicate with a wide range of audiences. Thanks, Lydia, for the opportunity learn and grow as a researcher in your lab.

I thank my committee members Hal Alper, George Georgiou, and Adron Harris for their investment of time, excellent questions, and valuable feedback on this work, pointing me in the most impactful directions.

All the members of my lab have contributed to my work in some way and are true friends. I especially thank Seung Hee for helping me get started and Grant, Steve, Jorge, Chen, and Kevin for their generosity with time and advice. My subgroup, Guided Investigations in Regulations of Large Systems (GIRLS), contributed a great deal of insight; thanks Seung Hee, Abby, Mia, Matsuri, Alex, and Rachael. I enjoyed daily exchange of ideas on science and life with my desk neighbor and running trainer Jo. I thank Mia for my office mug and support of teatime, Abby for her acronym-crafting skills, Angela for my origami aquarium and plant, and Mark for giving me a northern blot refresher course.

I enjoyed collaborating with coauthors on these projects. Shihui (Shane) Yang brought great wisdom to this work with his experience in *Zymomonas.* His lab members Wei Shen and Runxia Li helped characterize the Hfq 5'UTR in Chapter 2 and bring the paper to its published form. Cameron Blome performed Western blots and flow cytometry for Chapter 2. Juan Gonzalez-Rivera took on the challenge of learning RNA-RNA EMSAs for Chapter 3. Bobi Simonsen became a true teammate and played a major role in cloning and screening *Zymomonas* strains for Chapter 3. I am especially thankful for her dedication and enthusiasm in this work throughout her undergraduate career. Meg Arnold and Paige Williams brought much needed computational expertise to Chapter 4 and were so patient with the "translation" between biology and code. My first undergraduate assistant Emily Schmitz eagerly helped get the Chapter 3 project started. Our high school summer scholar Monona Khare caught on quickly and performed initial screens of strains in Chapter 4.

The McKetta Department of Chemical Engineering staff facilitated this work in a myriad of ways from the protection of my data from computer motherboard failure to the construction of shields to protect us from our radioactive samples.

Resources at the University of Texas at Austin enabled and accelerated this work: the Texas Advanced Computing Center, the Proteomics Facility, the Genomic Sequencing and Analysis Facility, the Bioinformatics Team and Byte Club, Library Services, and Environmental Health and Safety.

Financial support of my work came from the National Science Foundation Graduate Research Fellowship Program (DGE-1610403), the John and Mary Booker Endowed Graduate Fellowship, the Thrust 2000 – Meason/Klaerner Endowed Graduate Fellowship in Engineering, and the Dean's Prestigious Fellowship Supplement. Project funding came from the Welch Foundation (F-1756 to LMC), the Cancer Prevention Research Initiative

of Texas (RP110782), and the National Science Foundation CAREER Program (CBET-1254754 to LMC).

In addition to my research colleagues, I enjoyed personal support from many friends. To all those who upheld me personally and through prayer, thank you! I thank all my Missional Communities for constant encouragement and prayer, visits to the lab, labeling hundreds of Eppendorf tubes, and flexibility with my schedule. My roommates over the years made home peaceful and fun: Melissa, Esther, Chrissie, Cat, Aytilla, Christina, Ryn, Summer, Jenny, Jenn, and Gracey.

I would like to thank my family for their support over the years, especially my dad for going the extra mile(s) with science fair projects to put me on this trajectory and my mom for the encouraging cards she sends that keep me going.

Above all, I thank God for orchestrating these people and adventures with infinite wisdom and love. Cultivating bacteria humbled me and taught me a lot about you – your patience, care, ingenuity, and design for multiplication. To discover anything previously known only to you is a tremendous gift I will never forget. *For from Him and through Him and to Him are all things. To Him be the glory forever!*

Abstract

Uncovering and engineering complex phenotypes in bacteria using regulatory noncoding RNAs

Katherine Elizabeth Haning, PhD

The University of Texas at Austin, 2018

Supervisor: Lydia Contreras

Microorganisms exhibit amazing capabilities to sustainably produce biofuels, pharmaceuticals, and other chemicals. However, industrial stresses hinder efficient production. Metabolic engineering aims to overcome this challenge by genetically enhancing the metabolism of the organism, but existing approaches fail to fully control the complex cellular responses to stress, which are dynamic and involve large sets of genes. One untapped resource is the pool of native regulators known as noncoding RNAs that instinctively coordinate stress responses. Despite widespread discovery noncoding RNAs, few strategies harness their power to meet industrial goals. Most noncoding RNA functions remain unknown and lack foreseeable roles in producing phenotypes of interest. Moreover, existing methods to uncover noncoding RNA functions are slow and aimless, requiring laborious biochemical experiments and resulting in hit-or-miss relevance to any engineering goal.

To address this, we developed computational methods to efficiently identify noncoding RNAs relevant to a target phenotype. In the context of local, single-gene RNA regulators, we developed a strategy to identify 5' untranslated regions from transcriptome

data and to rapidly screen their responses to stress in vivo. With this discovery pipeline, we exposed the first 5' untranslated regions in ethanol-producing bacterium *Zymomonas mobilis* and illuminated their roles in stress responses. We also uncovered a network of multi-gene noncoding RNA regulators involved in the natural ethanol tolerance of *Z. mobilis*. To enable rapid identification of multi-gene noncoding RNA regulators like these in other organisms, we developed a computational approach to mine transcriptome data for noncoding RNAs with greatest potential impact on a phenotype of interest. Taken together, this work accelerates complex phenotype engineering in a variety of organisms by expanding the set of regulators available for engineering to include noncoding RNAs.

Table of Contents

List of Tables xiii

List of Figures xiv

Chapter 1: Introduction 1

1.1 Engineering Complex Phenotypes in Bacteria 1

1.2 Noncoding RNAs: Natural Regulators of Complex Phenotypes 2

1.3 Exploiting Noncoding RNAs for Engineering Goals 6

1.4 *Zymomonas* as a Model System 9

Chapter 2: Identification and Characterization of 5' Untranslated Regions (5'UTRs) in *Zymomonas mobilis* 11

2.1 Chapter Summary 11

2.2 Introduction 12

2.3 Results 14

2.3.1 Identification of 101 Potential 5'UTRs in *Z. mobilis* Using Transcriptomic Data 14

2.3.2 Discerning Properties of 5'UTRs through Bioinformatics Analysis 15

2.3.3 Validation of 36 5'UTR Candidates by RT-PCR Analysis 20

2.3.4 Establishment and Validation of the High-Throughput Fluorescence-Based Screening System 23

2.3.4 Identification and Characterization of Stress-Responsive Regulatory 5'UTRs Using Reporter Gene System 26

2.3.5 Native Regulation of Hfq Translation by 5'UTR 30

2.4 Discussion 34

Chapter 3: Small Noncoding RNAs Involved in Ethanol Tolerance in *Zymomonas mobilis* 39

3.1 Chapter Summary 39

3.2 Introduction....39

3.3 Results....42

3.3.1 sRNAs Affect Ethanol Tolerance and Production....42

3.3.2 Multi-omics Analyses Reveal Gene Networks Associated with Zms4 and Zms6....46

3.3.3 Identification of Potential Direct Targets of Zms4 and Zms6....50

3.3.4 Validation of Direct Targets by EMSA....52

3.3.5 Combinatorial Effects of sRNAs on Ethanol Tolerance Shows Complex Effect....61

3.4 Discussion....62

Chapter 4: A Bioinformatics Pipeline Developed to Identify Relevant Regulatory Small RNAs for Synthetic Phenotype Construction....68

4.1 Chapter Summary....68

4.2 Introduction....68

4.3 Results....73

4.3.1 Development of REFINE Approach to Identify sRNAs that Affect Phenotypes of Interest....73

4.3.2 REFINE Approach Identifies Well-characterized Iron Stress Regulator sRNA RyhB in *E. coli*....77

4.3.3 Identification of sRNAs in *Z. mobilis* that Improve Ethanol Tolerance....80

4.4 Discussion....87

Chapter 5: Major Findings and Perspectives....91

Chapter 6: Materials and Methods....96

6.1 Methods for Chapter 2....96

6.1.1 Bacterial Strain and Culture Conditions....96

6.1.2 Construction of *hfq* Mutant Strains....96

6.1.3 5' Rapid Amplification of cDNA Ends (RACE)....97

6.1.4 Construction of GFP-Reporter Plasmids with 5'UTRs....98

6.1.5 Fluorescence Measurements....98

6.1.6 Quantitative RT-PCR....99

6.1.7 Western Blotting Analysis and Quantification of Protein Expression Levels....100

6.2 Methods for Chapter 3....101

6.2.1 Strains and Culture Conditions....101

6.2.2 Ethanol Assay....103

6.2.3 RNA Purification....103

6.2.4 Protein Purification....104

6.2.5 RNA Sequencing and Analysis....104

6.2.6 Mass Spectrometry....105

6.2.7 MS2-affinity Purification of sRNAs in vivo....105

6.2.8 PCR Amplification and in vitro Transcription of RNAs for Binding Assays....107

6.2.9 Detection of RNA-RNA Interactions....108

6.2.10 Northern Blot Analysis....109

6.3 Methods for Chapter 4....110

6.3.1 Calculating Intergenic Region Coverage from Existing RNA-seq Data....110

6.3.2 sRNA Prediction with sRNAscout....110

6.3.3 Scoring sRNA Candidates for Phenotype Effect with sRNAphenoscore....111

6.3.4 sRNA Candidate Overexpression Strain Development 111
6.3.5 Evaluating Strain Performance 112
Appendices.....113
Appendix A: Supplementary Tables.....113
Appendix B: Supplementary Figures.....201
Appendix C: Software Developed in This Work.....204
REFINE (Chapter 4) 204
User Guide 204
files.R.....207
func2_dplyr.R 207
toolbox.R.....214
run_DESeq2.R 220
References.....222

List of Tables

Table 2.1: List of 5'UTR candidates with features. 18

Table 3.1: List of transcripts and proteins differentially regulated by both ethanol stress and sRNA overexpression in *Z. mobilis* 49

Table 3.2: Confirmed targets of Zms4 and Zms6 regulation. 55

Table 4.1: Filtering criteria used by sRNAscout to identify sRNA candidate peak regions in each dataset. 83

Table 4.2: Phenoscores from aerobic vs. anaerobic data of previously discovered *Z. mobilis* sRNAs. 83

Table A.1: List of 101 initial 5'UTR candidates and their features. 113

Table A.2: List of final 5'UTR candidates and their features. 119

Table A.3: List of primers for 5'UTR study. 121

Table A.4: Transcripts significantly dependent on Zms4 and Zms6 140

Table A.5: Differentially expressed proteins upon Zms4 and Zms6 induction. 153

Table A.6: Plasmids and primers for *Z. mobilis* sRNA studies. 168

Table A.7: Transcripts that co-immunoprecipitate with Zms4 and Zms6. 172

Table A.8: Proteins that co-immunoprecipitate with Zms4 and Zms6. 179

Table A.9: sRNA candidate regions for *Z. mobilis* oxygen stress. 181

List of Figures

Figure 1.1: ncRNA search and characterization. 4
Figure 1.2: Two classes of ncRNA regulators. 5
Figure 2.1: Pipeline for the selection of 5′UTR candidates. 15
Figure 2.2: Structural analysis of UTR candidates using LocARNA. 17
Figure 2.3: RT-PCR analysis to experimentally confirm 5′UTR candidates. 21
Figure 2.4: Experimental analysis of 5′UTRs by RT-PCR. 22
Figure 2.5: Summary of the results of 5′ RACE. 22
Figure 2.6: Establishment of a GFP-based high-throughput reporter gene screening system to characterize regulatory 5′UTR regions. 25
Figure 2.7: Identification of 5′UTRs that regulate GFP expression under ethanol stress. 28
Figure 2.8: The effect of acetate (A) and xylose (B) on 5′UTRs. 29
Figure 2.9: Growth curves of the wild type and ΔUTR_ZMO0347 strains under a range of ethanol stresses. 31
Figure 2.10: Native regulation of Hfq in ethanol stress. 33
Figure 2.11: Summary of observed Hfq regulation in ethanol stress. 37
Figure 3.1: Verification of transcript levels in sRNA overexpression strains. 44
Figure 3.2: Verification of transcript levels in sRNA deletion strains. 44
Figure 3.3: sRNA expression levels affect ethanol tolerance and production. 45
Figure 3.4: Experimental approaches designed to discern gene networks associated with Zms4 and Zms6. 47
Figure 3.5: Integrated omics analyses reveal most likely targets of Zms4 and Zms6 regulation. 51

Figure 3.6: sRNA-target interacting pairs verified by EMSA. 53
Figure 3.7: Detection of Zms4 and Zms6 in vitro interactions with RNA targets by EMSA. 54
Figure 3.8: Identification of Zms4 and Zms6 binding sites with their mRNA targets. ... 57
Figure 3.9: Predicted binding sites for direct targets of Zms4 and Zms6 by IntaRNA. ... 59
Figure 3.10: Competition binding assay shows simultaneous target binding. 60
Figure 3.11: Combined sRNA overexpression indicates complexity of the network. 62
Figure 3.12: A network of sRNA-target interacting pairs were identified in this study. .. 65
Figure 4.1: REFINE approach identifies sRNAs relevant to phenotypes of industrial interest. 72
Figure 4.2: Computational workflow from raw RNA-seq data to phenoscores of each predicted sRNA candidate. 74
Figure 4.3: sRNAscout identifies well-characterized iron-starvation regulator RyhB from existing *E. coli* transcriptome data. 78
Figure 4.4: Novel sRNAs in *E. coli* discovered by sRNAscout. 80
Figure 4.5: sRNAscout identifies previously known *Z. mobilis* sRNAs responsive to oxygen stress. 82
Figure 4.6: Novel sRNAs in *Z. mobilis* discovered by sRNAscout. 85
Figure 4.7: *Z. mobilis* sRNAs identified by the REFINE process impact stress tolerance. 86
Figure B.1: Flow cytometry histograms for screening all UTR-GFP construct responses to stress. 201

Chapter 1: Introduction

1.1 Engineering Complex Phenotypes in Bacteria

As cellular factories, microorganisms produce biofuels, pharmaceuticals, and other chemicals in sustainable ways. A main hurdle to optimizing strains has been our inability to fully control cellular phenotypes to achieve high yields. The industrial process presents a number of challenges to any biocatalyst. In addition to the stress of toxic end products (like ethanol) produced during fermentation, the pretreatment of biomass to release sugar monomers also releases inhibitory compounds that cripple cell growth, metabolism, and production [1]. These inhibitors include furfural, hydroxymethylfurfural (HMF), weak acids such as formic acid, levulinic acids, acetic acid, phenolic aldehydes, and inorganic salts [1]. Stress tolerance is a complex, multigenic trait that cannot elicited by classic single-gene metabolic engineering [2].

So far, complex phenotype engineering involves long-course adaptation experiments with random evolution by chemical mutagenesis [3]. It is difficult to transfer these evolved capabilities to other strains and requires screening of large combinatorial strain libraries. Systems biology efforts focus on characterization of stress tolerance networks through transcriptome, proteome, and metabolome experiments. Potential regulators may be inferred from these studies, but ultimately, slow biochemical assays are used to verify each interaction and finally inform a strain engineering strategy. Although we continue to gather more and more information about the integrated gene networks behind stress tolerance, we lack rational strategies to produce these complex phenotypes on demand.

In this work, I investigate how to manipulate global regulatory pathways to produce cellular phenotypes of interest. How can we locate natural, dynamic regulators and characterize their networks in high-throughput ways? What complex phenotypes can be achieved by synthetically tuning these regulators? This work presents steps forward in these questions.

1.2 Noncoding RNAs: Natural Regulators of Complex Phenotypes

Recently, noncoding RNAs (ncRNAs) were discovered as central molecules in global regulatory pathways, highly efficient at coordinating survival when cells experience environmental stresses [4–6]. These RNA transcripts carry the name "noncoding" because of their discovery in previously unannotated regions of the genome and because they function without being translated to protein (although some do encode small peptides). An advantage of ncRNAs compared to protein regulators like transcription factors is this efficiency. Because they do not require translational machinery to be made and because they act directly on existing mRNA transcripts, ncRNAs exert their control rapidly [7]. This dynamic nature and low metabolic burden make ncRNAs especially suitable to coordinate stress responses including temperature, nutrient, membrane, oxidative, iron, pH, and anaerobic stresses [4,5].

In *E. coli* alone, over 100 ncRNAs are now confirmed. Many ncRNAs impact expression of multiple mRNA targets, some of which have been biochemically and genetically identified and others remain to be uncovered. A variety of ncRNAs are conserved across many organisms, including pathogenic bacteria, which depend on complex stress adaptation mechanisms to invade and thrive within their hosts [8].

Following the surge of ncRNA discovery (largely enabled by high-throughput sequencing), investigations in *E. coli* and other model organisms shifted toward

mechanistic studies. **Figure 1.1** summarizes the large-scale sRNA searches, validation experiments, mechanistic and functional studies, and ultimately, targeting efforts. Through mechanistic studies, researchers observed regulatory impacts of ncRNAs on mRNA and protein expression. Two main classes define regulatory ncRNAs by their genomic locations relative to their targets: *cis*-acting and *trans*-acting (**Figure 1.2**). *Cis*-acting ncRNAs are encoded as part of the same transcript as their target mRNA, while *trans*-acting ncRNAs reside apart from their targets in the genome.

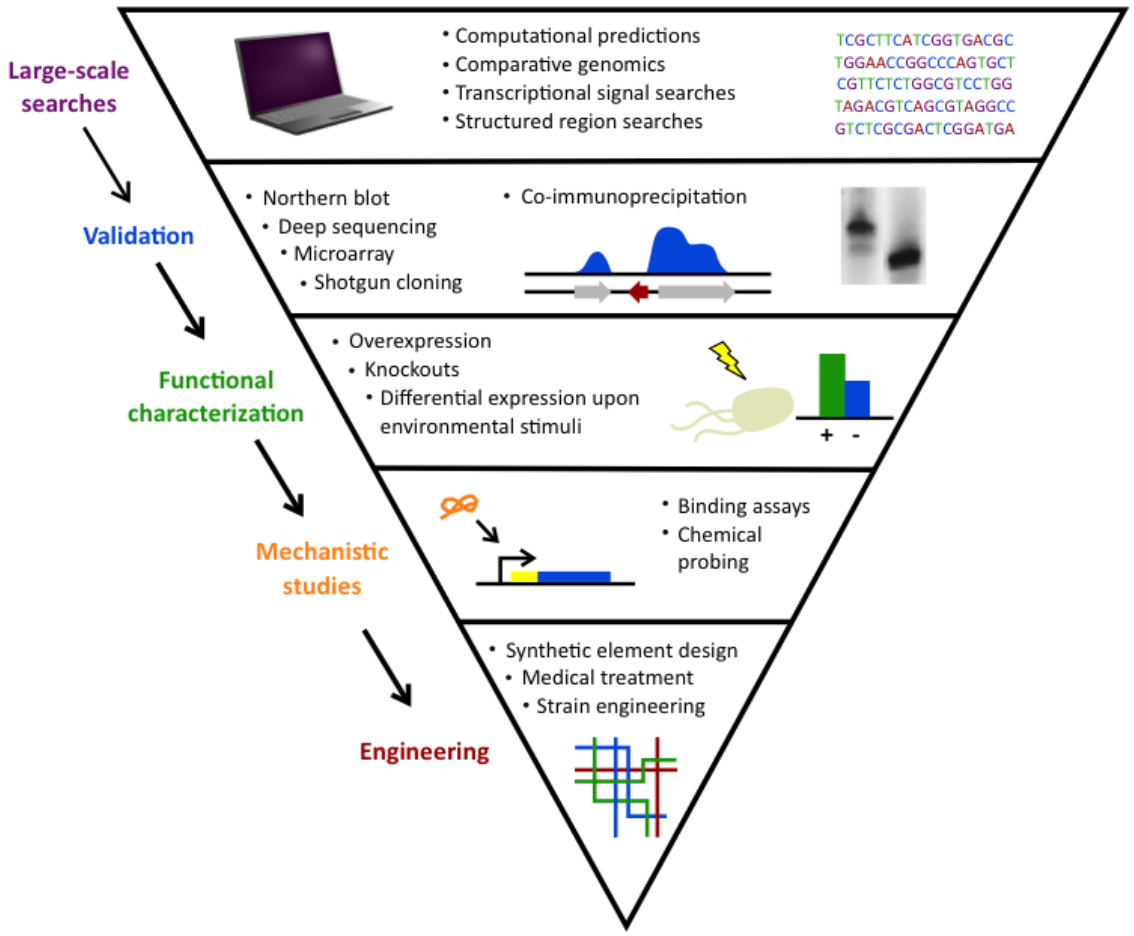


Figure 1.1: ncRNA search and characterization.

Discovery of ncRNAs often begins with large-scale computational searches followed by experimental validation. Functional characterization of confirmed candidates identifies their gene or protein targets and mechanistic studies elucidate their methods of action. Finally, ncRNAs can be used in engineering efforts to develop useful applications from synthetic elements to medical treatments.[9]

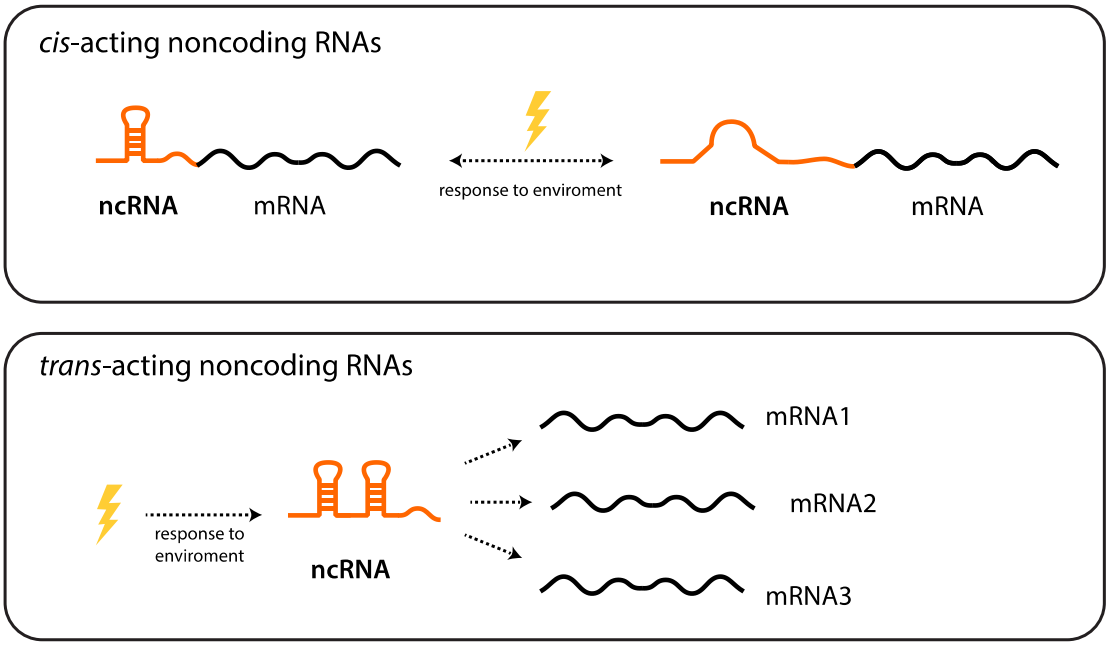


Figure 1.2: Two classes of ncRNA regulators.

ncRNAs are generally classified by their genomic location relative to their targets. *Cis*-acting ncRNAs are part of the mRNA transcript they regulate. *Cis*-acting ncRNAs respond directly to environmental signals with structural changes, thereby altering the stability and/or translation of the mRNA target. *Trans*-acting ncRNAs are independent transcripts from their mRNA targets. They respond in expression level to environmental or growth phase signals and regulate their targets through base-pairing interactions. *Trans*-acting mRNAs may have multiple mRNA targets of a variety of functions.

Cis-acting ncRNAs are natural biosensors, directly responsive to environmental changes including temperature, pH, and metabolites [10]. These RNA elements include riboswitches, a class of powerful regulators typically found in the 5' untranslated regions (5'UTRs) of their mRNA targets. As "switches," they are known for their conformational changes, which trigger shifts in downstream mRNA translation or stability. In this category, RNA thermometers respond structurally to changes in temperature to regulate translation [11].

Trans-encoded ncRNAs are often called "small RNAs" (sRNAs) considering their short 50–500 nt transcript length compared to most mRNAs. These sRNAs regulate mRNA and protein expression, typically through base-pairing interactions with mRNA targets. Upon binding, they block translation or change stability in ways that impact protein expression [12]. Some sRNAs (although fewer) regulate proteins directly [13]. A single sRNA may have dozens of direct mRNA and protein targets and thereby influence multiple cellular pathways. Each sRNA-mRNA interaction may produce a number of indirect effects as well. An example is the sRNA-regulation of a transcription factor protein, which has its own set of downstream targets [14]. Many *trans*-encoded sRNAs depend on Hfq, an RNA-binding protein, that provides stability to enable sRNAs to function, particularly in gram-negative bacteria [15]. Hfq and other RNA-binding proteins represent an additional layer of regulation in sRNA networks.

1.3 Exploiting Noncoding RNAs for Engineering Goals

With improved mechanistic understanding of ncRNAs came enormous interest to inhibit and/or mimic their natural functions to achieve desired changes in targeted genes and their networks [16,17]. Their natural regulatory roles highlight ncRNAs as powerful engineering tools [17–20].

Regulatory 5'UTRs present great potential for local control of gene expression in metabolic engineering efforts [21]. As *cis*-acting regulators, their impacts on their targets are extremely fast. Sequence and homology searches reveal broad conservation of many canonical riboswitches, but more unique classes remain to be uncovered [22]. Due to their short length and simple switch mechanism, synthetic riboswitches have been readily engineered as controllers of metabolic flux, sensors for screening or reporting, and chemically inducible expression systems [23]. In Chapter 2, I discuss the discovery and

characterization of natural regulatory 5'UTRs in ethanol-producing bacterium *Zymomonas mobilis*.

Most *trans*-acting ncRNA engineering efforts so far mimic ncRNA antisense binding to block translation of individual mRNA targets [24,25]. The short base-pairing interaction between a synthetic ncRNA and a single mRNA target can be easily designed [26]. However, a major asset of natural *trans*-ncRNAs remains largely untapped: the regulation of *multiple* mRNA targets to coordinate complex network responses. This stems from a lack of knowledge about their networks of mRNAs, proteins, and transcriptional factors that are regulated in response to environmental changes.

Recently, natural ncRNA engineering increased acid tolerance in *E. coli* [27]. These efforts were largely enabled by prior knowledge of acid resistance systems and the wide variety of genes involved. One acid resistance mechanism depends on the RpoS sigma factor that directly or indirectly regulates about 500 genes [28,29]. Given this dependence on RpoS, engineering efforts focused on manipulating RpoS levels. Interestingly, simple overexpression of *rpoS* by removal of its natural 5'UTR does not significantly improve acid tolerance and is not sufficient to fully induce the genes it regulates [30]. The inability to use conventional overexpression strategies to increase levels of RpoS is not surprising given the multi-layer regulation of this protein by a variety of ncRNAs (e.g. DsrA, RprA, and ArcZ) that stimulate *rpoS* translation through its 5'UTR [30]. This is an interesting case in which the *cis*-regulatory ncRNA (5'UTR) blocks the RBS until the *trans*-regulatory ncRNAs (DsrA, RprA, ArcZ) bind to complementary nucleotide sequences in the *cis*-regulator and free the RBS to allow translation.

A complex phenotype like acid tolerance is well suited for *trans*-regulatory ncRNA engineering. A new strategy overexpressed the three ncRNAs known to up-regulate RpoS levels (instead of overexpressing RpoS directly) [27]. All three overexpression strains (RprA,

ArcZ, and DsrA under their native promoters) showed increased survival percentage after one hour of growth in acidified media (pH 2.5) with three-fold, five-fold, and 106-fold improvements over the empty plasmid control strain, respectively. Furthermore, a triple overexpression strain (combination of all three ncRNAs) showed an impressive 8500-fold improvement in cell survival after acid stress, demonstrating a supra-additive effect of combinatorial overexpression that cannot be solely attributed to simple increase in RpoS protein expression levels.

Engineering ncRNAs to improve acid tolerance set a precedent for using natural ncRNA regulators to tune expression of entire sets of pathways. However, unlike this demonstration, most ncRNAs remain uncharacterized and lack known mRNA targets and mechanisms, particularly beyond *E. coli*. Furthermore, the way in which ncRNAs can be used as building units in synthetic tolerance networks remains unclear. In Chapter 3, I highlight the characterization of a network of ncRNAs and their targets in *Z. mobilis* that are anticipated to be useful in engineering ethanol tolerance.

In light of the low-throughput nature of mechanistic characterization of ncRNAs both in vitro and in vivo (and the difficulty of performing in vivo studies at all in some non-model organisms), I developed a *computational* method to efficiently identify ncRNAs relevant to a target phenotype. As described in Chapter 4, we anticipate this to be a useful tool for metabolic engineers to approach the large (and growing) pool of ncRNA regulators with an engineering goal in mind. Previously, this has been impractical, not because the ncRNAs necessarily lack power, but because they lack foreseeable roles in producing phenotypes of interest.

1.4 *ZYMOMONAS* AS A MODEL SYSTEM

As a model system for my work, I focus on *Z. mobilis.* This gram-negative bacterium naturally produces ethanol and exhibits many desirable industrial qualities [31–35]. *Z. mobilis* is generally regarded as safe (GRAS) and exhibits high ethanol tolerance up to 16% (v/v) [36]. It produces ethanol across a broad pH range (3.5–7.5, especially low pH) [37]. As a facultative anaerobe, *Z. mobilis* does not require controlled aeration during fermentation, reducing production costs [35]. *Z. mobilis* ferments glucose to ethanol with the efficient Entner-Doudoroff (ED) pathway [36,38]. Compared to the classical Embden-Meyerhof-Parnas (EMP) pathway for glycolysis in other model species such as *Saccharomyces cerevisiae* and *E. coli*, the ED pathway yields only one mole of ATP per mole of glucose [38].

Multiple approaches have been explored to understand and engineer tolerance to stresses in *Z. mobilis* [39]. Over the last few decades, metabolic engineering and directed evolution developed a variety of strains with unique tolerance capabilities to ethanol, acid, sodium acetate, salt, furfural, and vanillin [40–46]. Complete genome sequences are available for several strains, [47–49] as well as transcriptome and proteome profiles under ethanol, furfural, acetate, and anaerobic stress conditions [48,50–53]. However, on their own, genome-wide omics studies have limited impact in guiding strain engineering since they only yield correlations between large sets of genes and specific strain and growth conditions.

One link between large sets of differentially expressed genes and specific implementable strategies is the identification of regulators that coordinate these large sets. A single regulator (or a small set) can be modified to shift the network and induce improved phenotypes. Global transcription machinery engineering illustrates this concept. Applied to *Z. mobilis*, global transcription machinery engineering improved ethanol tolerance by the mutation of transcription factor RpoD (σ^{70}) [40,54].

As described before, ncRNAs efficiently regulate networks of genes and have now been discovered in organisms of traditionally high relevance to biotechnology including *Bacillus*, *Clostridium*, and *Synechocystis* [55–57]. Recent discovery of 15 ncRNAs in *Z. mobilis* with differential expression under high and low ethanol conditions inspires us to consider the potential of natural ncRNA regulation of stress tolerance in this organism [53]. We envision ncRNA regulators with a variety of mechanisms and targets that can be discovered and employed for engineering

Engineering extremophiles such as *Z. mobilis* can be more challenging than model organisms because of the lack of established genetic tools and protocols to accomplish desired phenotypes. But, a fruit of this challenge has been more innovative approaches to find impactful ncRNA regulators while minimizing the experimental effort of genomic manipulation and screening. A goal of this work is to translate fundamental understanding of *Z. mobilis* tolerance mechanisms into strategies for strain engineering, both in *Z. mobilis* and in other organisms.

Chapter 2: Identification and Characterization of 5' Untranslated Regions (5'UTRs) in *Zymomonas mobilis*

2.1 CHAPTER SUMMARY

Regulatory RNA regions within a transcript, particularly in the 5' untranslated region (5'UTR), have been shown in a variety of organisms to control the expression levels of these mRNAs in response to various metabolites or environmental conditions. Considering the unique tolerance of *Zymomonas mobilis* to ethanol and the growing interest in engineering microbial strains with enhanced tolerance to industrial inhibitors, we searched natural *cis*-regulatory regions in this microorganism using transcriptomic data and bioinformatics analysis. Potential regulatory 5'UTRs were identified and filtered based on length, gene function, relative gene counts, and conservation in other organisms. An *in vivo* fluorescence-based screening system was developed to confirm the responsiveness of 36 5'UTR candidates to ethanol, acetate, and xylose stresses. UTR_ZMO0347 (5'UTR of gene ZMO0347 encoding the RNA binding protein Hfq) was found to down-regulate downstream gene expression under ethanol stress. Genomic deletion of UTR_ZMO0347 led to a general decrease of *hfq* expression at the transcript level and increased sensitivity for observed changes in Hfq expression at the protein level. The role of UTR_ZMO0347 and other 5'UTRs gives us insight into the regulatory network of *Z. mobilis* in response to stress and unlocks new strategies for engineering robust industrial strains as well as for harvesting novel responsive regulatory biological parts for controllable gene expression platforms in this organism. *

* This chapter has been previously published.[162]

2.2 INTRODUCTION

Biomass pretreatment and hydrolysis releases sugar monomers from cellulose and hemicellulose. Several growth inhibitors are released from this process including xylose and acetate [58,59]. *Zymomonas mobilis* is a promising ethanologenic bacterium due to its efficient ethanol production and high ethanol tolerance (16% v/v). Several recent reviews have outlined the progress that has been made in understanding ethanol-related pathways in this organism [32,34,35]. Originally, wild-type *Z. mobilis* could only utilize glucose, sucrose, and fructose as carbon sources for ethanol production, but metabolic engineering approaches have enabled xylose and arabinose metabolisms [60–62].

Transcriptomic and proteomic analyses of *Z. mobilis* in ethanol-supplemented conditions revealed that genes associated with DNA repair, membrane biogenesis, carbohydrate metabolism, transport, and transcriptional regulation are differentially expressed in ethanol stress, showing the complexity of this phenotype [52,63]. Additional omics studies have shown both xylose and acetate as important inhibitory factors of *Z. mobilis* growth and ethanol production [64]. Upon co-utilization of xylose and glucose, especially with inhibitors such as acetate or furfural, the gene expression of redox mechanisms and carbon and energy metabolisms are dramatically changed [50,64]. While it is well-documented that acetate toxicity negatively affects cell growth and ethanol production [64,65], the direct underlying molecular mechanisms involved in stress responses and non-natural sugar utilization still need to be further explored. To uncover potential stress response mechanisms in response to various inhibitors in *Z. mobilis*, we focused on the discovery of regulatory RNAs using transcriptomic data and bioinformatics analysis [53]. Fifteen small RNAs (sRNAs) were previously identified suggesting the potential existence of other types of regulatory non-coding RNAs in this organism, such as regulatory 5'UTRs that have not yet been annotated [53].

Regulatory RNAs include 5' and 3' untranslated regions (UTRs), riboswitches, *cis*-acting antisense RNAs, and *trans*-acting small non-coding RNAs that regulate gene expression, sometimes in response to external stress [17,53,66]. 5'UTRs have been reported to modify gene regulation in both prokaryotes and eukaryotes on the basis of the changes in temperature, pH, and other metabolites [10]. For example, the 5'UTR of *prfA* mRNA immediately responds with a structural change upon temperature changes in *Listeria monocytogenes*, which is critical for survival for pathogenic bacteria in the host [67]. The 5'UTR of *alx* gene is a pH sensor in *Escherichia coli* that changes structure to allow translation of *alx* in alkaline conditions [68].

Riboswitches represent yet another class of sensors in which metabolites control gene expression in various metabolic pathways. Upon sensing small molecule metabolites, riboswitches trigger structural changes to regulate transcription or translation of mRNAs. Riboswitches consist of two components: an aptamer and an expression platform. Aptamer domains are between 35 and 200 nucleotides and responsible for direct binding to small molecule metabolites such as ions, nucleotides, amino acids, or coenzymes [69–71]. While aptamer domains are highly structured and conserved among different species, expression platforms can vary in sequence and structure and undergo different conformational changes in response to ligand binding to aptamer domains, resulting in altered downstream gene expression, either activation or repression [17,72–74].

Interestingly, a new class of RNA elements, OLE (ornate, long and extremophile), is highly expressed, stable, and interact with OLE-associated protein (OAP) to protect extremophiles in response to ethanol stress [75,76]. Although there is no evidence of OLE RNAs in *Z. mobilis*, this recent finding provided us with the hypothesis that ethanol-responsive RNA elements could be present throughout the *Z. mobilis* genome and that these could be important for its ethanol tolerance capability.

Although many different types of regulatory elements such as riboswitches and 5'UTRs have been discovered among various bacterial species, none have been experimentally confirmed in *Z. mobilis*. Three riboswitches (two cobalamine and one TPP riboswitches) have been predicted in *Z. mobilis* genome by computational analysis using CMfinder, suggesting that these regulatory elements are likely present in this organism [77]. In this work, available transcriptomic data were searched with a bioinformatics approach to discover potential 5'UTRs that play a role in ethanol stress response. These were then experimentally tested for their regulatory roles and biological relevance under stress conditions utilizing an *in vivo* GFP (Green Fluorescence Protein) reporter system and classical genetic approaches.

2.3 RESULTS

2.3.1 Identification of 101 Potential 5'UTRs in *Z. mobilis* Using Transcriptomic Data

Using previously published datasets [53], transcripts were screened for their expression in the 5'UTR regions of the adjacent coding regions to discover putative 5'UTR regions that could contribute to the gene regulation in *Z. mobilis*. Initially, 392 potential candidates were identified, leading to the need of filtering these 5'UTR candidates from the large number and diversity of other types of transcripts. For this purpose, a bioinformatics pipeline was developed to select candidates from the large number of initial candidates (**Figure 2.1**). Only candidates that showed comparable levels of expression with the adjacent genes were retained as potential UTRs. All transcripts <35 base pairs were filtered out, as this is the shortest known length of a UTR regulatory element [71]. As well-characterized metabolic enzymes are highly regulated by their 5' UTRs, these types of candidates were prioritized in the list. Collectively, this analysis resulted in a total of 101 potential candidates that were selected for experimental analysis (**Table A.1**).

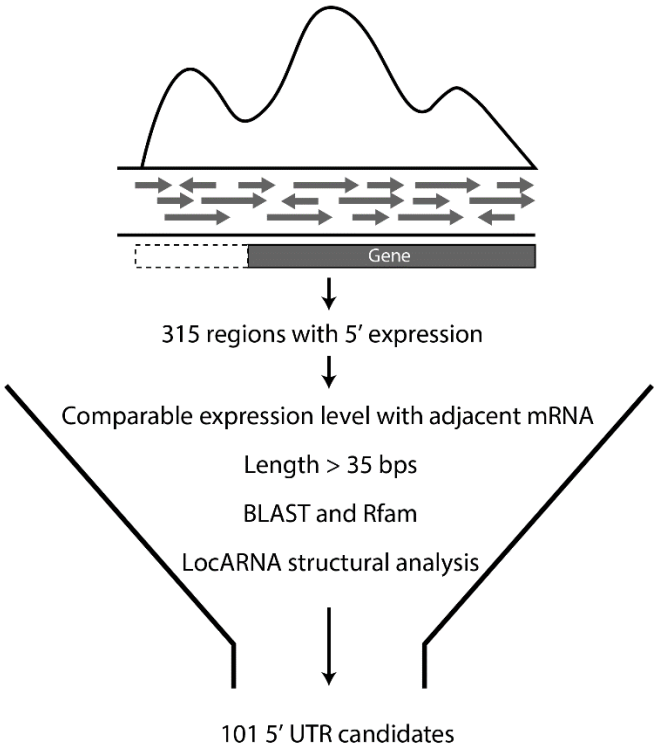


Figure 2.1: Pipeline for the selection of 5′UTR candidates.

Initial 5′UTR candidates were selected from transcriptomic data and then filtered for comparable expression level to adjacent mRNA, with length more than 35 bp. 5′UTR candidates that met these criteria were analyzed for homology with known RNA regulators by Rfam and for general conservation in other organisms by BLAST. From this analysis, 101 candidates were selected for experimental confirmation.

2.3.2 Discerning Properties of 5'UTRs through Bioinformatics Analysis

To further discern the properties of potential 5'UTR candidates, bioinformatics analyses were performed using Rfam (a database of multiple sequence alignments), consensus secondary structures, and covariance models representing RNA families [78]. There are three RNA categories in Rfam: non-coding RNAs, structured *cis*-regulatory elements, and self-splicing RNAs [79]. Four candidates were matched with predicted riboswitches in *Z. mobilis*: TPP, crcB, and two cobalamin switches (**Table 2.1**). TPP and

cobalamin riboswitches are widely conserved and have been demonstrated in *E. coli* and other bacteria, such as *Bacillus subtilis*, to control the regulation of downstream genes by direct binding to thiamine pyrophosphate and cobalamine, respectively [72,80]. The crcB RNA motif identified by Rfam is known as a fluoride riboswitch [81] adjacent to the 5' end of the chloride channel protein gene (ZMO0547). This fluoride switch regulates gene expression based on its structural change in response to fluoride ions. Genes encoding this fluoride-specific type of chloride channel protein have been shown to be regulated by fluoride riboswitches in a variety of organisms [82]. Importantly, these results supported the presence of 5'UTRs before experimental confirmation and also validated the UTR bioinformatics prediction methods.

Given that structural conservation is closely associated with the regulatory roles of RNA [83], LocARNA was used to determine conservation based on both sequence and structure [84]. NCBI BLAST [85] was used to identify sequence homology for each 5'UTR candidate, as required for LocARNA input. Of the initial 101 5'UTR candidates, 28 contained structurally conserved motifs, many of which showed complexity. **Figure 2.2** shows representative data of this analysis.

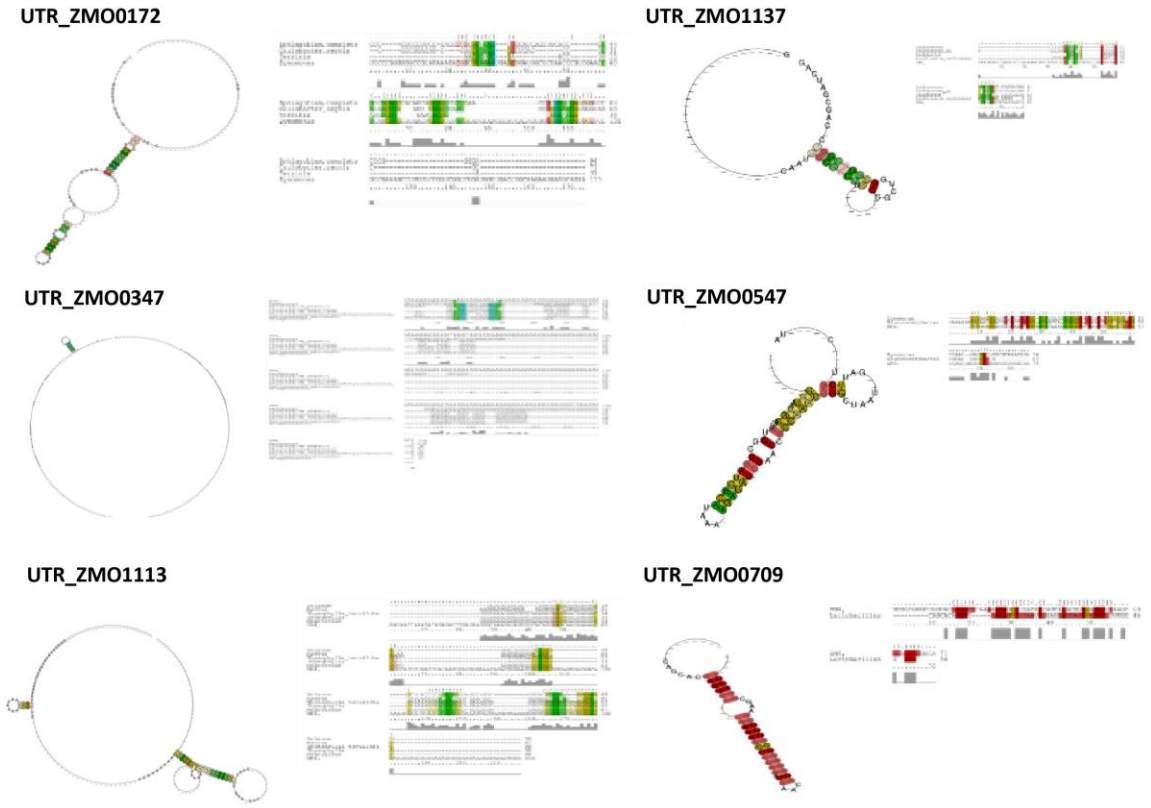


Figure 2.2: Structural analysis of UTR candidates using LocARNA.

Lastly, differential expression data under ethanol, acetate, and xylose stresses were compiled for the genes associated with the 5'UTR candidates [51,52,63,64]. Genes with differential expression under one or more stresses may be regulated by a 5'UTR or other mechanism. As shown in **Table 2.1** and **Table A.3**, 17 mRNAs and 7 proteins corresponding to the 5'UTR candidates were up- or down- regulated under stress. Due to their association with stress response(s), these candidates were further investigated in this work.

Table 2.1: List of 5'UTR candidates with features.

Gene ID	Gene	Conservation	Rfam	Other organisms
ZMO0172	Thiamin biosynthesis protein/phosphomethylpyrimidine synthase	Highly conserved	TPP	TPP riboswitch in *E. coli*
ZMO0979	TonB-dependent receptor	Highly conserved	Cobalamin	AdoCbl
ZMO1000	5-methyltetrahydropteroyltriglutamate	Highly conserved	Cobalamin	metE in *E. coli* and *Bacillus*
ZMO0547	chloride channel core	Conserved with *Gluconobacter*	*crcB*	Fluoride riboswitches
ZMO0056	glucosamine/fructose-6-phosphate aminotransferase	-	-	*glmS* ribozyme
ZMO0376	ATP-dependent protease La	-	-	-
ZMO0546	sulphate transporter	-	-	ABC transporter family
ZMO0660	dnaK molecular chaperone DnaK	-	-	-
ZMO1069	molecular chaperone DnaJ	Highly conserved	-	-
ZMO1139	acetolactate synthase large subunit	-	-	-
ZMO1142	thioredoxin reductase	-	-	-
ZMO0709	phosphoribosylaminoimidazole synthetase	Conserved with *Lactobacillus*	-	-
ZMO1137	phosphoserine phosphatase SerB	Somewhat conserved	-	*serC* regulated by glycine riboswitch
ZMO0187	3-deoxy-7-phosphoheptulonate synthase	-	-	-
ZMO0369	glucokinase	-	-	-

Table 2.1: List of 5'UTR candidates with features. (continued)

ZMO0405	ATP-dependent Clp protease ATP-binding subunit ClpA	-	-	-
ZMO0937	aromatic amino acid aminotransferase	-	-	-
ZMO1179	(uracil-5)-methyltransferase	-	-	-
ZMO1275	threonine dehydratase	Conserved with *Clostridium*	-	-
ZMO0689	oxidoreductase domain-containing protein	-	-	-
ZMO0131	metallophosphoesterase	-	-	-
ZMO0748	cysteine synthase	-	-	-
ZMO1034	calcium-binding EF-hand-containing protein	-	-	-
ZMO1113	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	-	-	-
ZMO0347	RNA-binding protein Hfq	Conserved with *Clostridium*	-	-
ZMO1198	5-aminolevulinate synthase	-	-	-
ZMO1478	6-phosphogluconolactonase	-	-	-
ZMO0275	ABC transporter	-	-	-
ZMO1399	fatty acid hydroxylase	-	-	-
ZMO1432	fusaric acid resistance protein	-	-	-
ZMO0367	glucose-6-phosphate 1-dehydrogenase	-	-	-
ZMO1412	MucR family transcriptional regulator	-	-	-
ZMO1048	phosphate ABC transporter inner membrane subunit PstC	-	-	-
ZMO0140	protein tyrosine phosphatase	-	-	-
ZMO0366	sugar transporter	-	-	-
ZMO1612	toluene tolerance family protein	-	-	-

2.3.3 Validation of 36 5'UTR Candidates by RT-PCR Analysis

The cellular expression of all 101 final candidates (**Table A.1**) was further confirmed by RT-PCR analysis, for which two primer sets were designed (**Figure 2.3A**). Primer set A was designed for the amplification of a long transcript from hypothetical 5'UTR regions to the middle of the corresponding mRNA coding region. Primer set B was designed to amplify a relatively short transcript inside the adjacent mRNA coding region as a positive control representing the adjacent mRNA expression level (**Figure 2.3A**). As a negative control, reverse transcriptase was excluded from the reaction to confirm the absence of genomic DNA contamination. Representative data are illustrated in **Figure 2.3B** and all RT-PCR results are shown in **Figure 2.4**. PCR bands from both primer sets A and B proved the expression of transcript containing potential UTRs upstream. However, the presence of a band from primer set B and not from set A, indicated that the UTR was not detected upstream of the mRNA. From the experimental analysis, fifty 5'UTR candidates that showed contiguous expression in the transcript along with the mRNA, which are marked as confirmed candidates in **Table A.1**. Lack of detectable expression of other candidates could be explained by unsuccessful PCRs, false positives in our candidate selection from the transcriptomic data, or different growth conditions between RT-PCR and previous transcriptomic studies.

Precise transcription start sites (TSS) for all candidates of interest were determined by 5' RACE (**Figure 2.3C**). A summary of the 36 final candidates that were further investigated and their gene function is included in **Table 2.1** and additional information in **Table A.2**. It should be noted that 14 candidates were excluded from further analysis out of the 50 experimentally confirmed candidates due to their overlapping transcription with the adjacent gene (3' end of the adjacent gene was connected with potential 5' UTRs), length <30 bp, or lack of experimental validation by RT-PCR and/or 5'RACE (**Figure 2.5**).

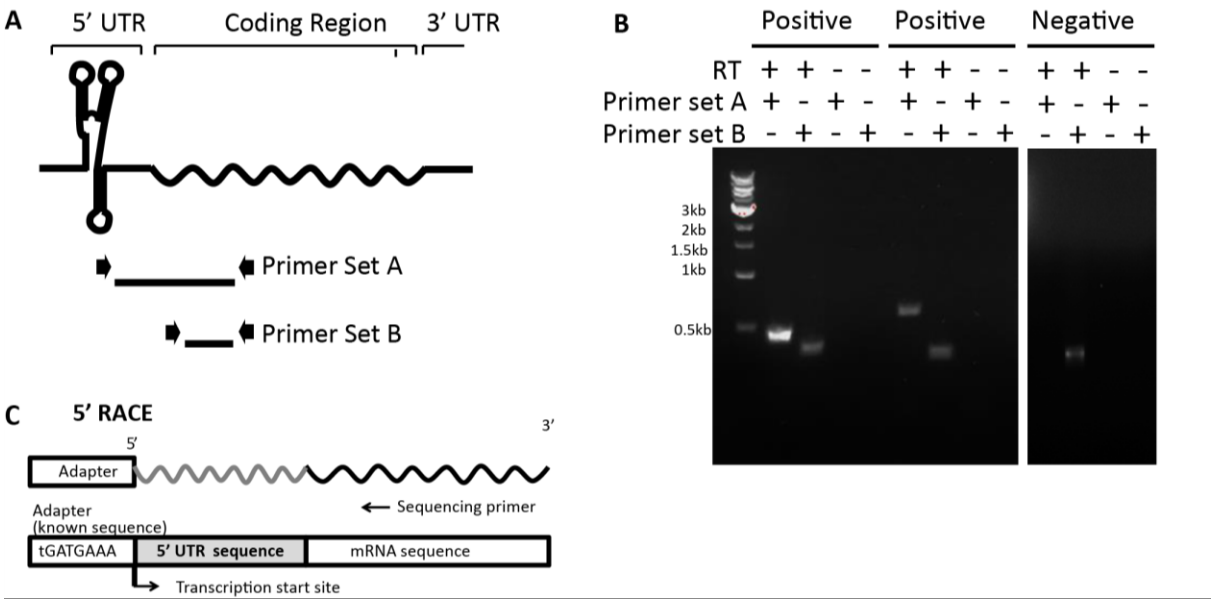


Figure 2.3: RT-PCR analysis to experimentally confirm 5′UTR candidates.

(A) To confirm the 5′UTRs detected in transcriptomic data are actually transcribed with their downstream mRNAs, two sets of primers were designed. Primer Set A amplifies from the middle of the predicted UTR into the adjacent mRNA coding region. Primer Set B, as a control, amplifies the coding region only, indicating the amplification level of the transcript. **(B)** Examples of positive (confirmed) and negative (undetected) 5′UTR candidates are shown. RNA was reverse-transcribed and then subjected to PCR with Primer Set A and Primer Set B for each 5′UTR candidate. As a negative control, RNA was left without reverse transcription to indicate background levels of any residual DNA that could be left undigested from DNaseI treatment. Samples were then visualized on agarose gels, as shown in the examples. **(C)** For each 5′UTR confirmed by RT-PCR, 5′RACE was performed to determine the transcription start site by attaching an adapter of known sequence to the 5′ end and sequencing from the mRNA coding region upstream toward the adapter.

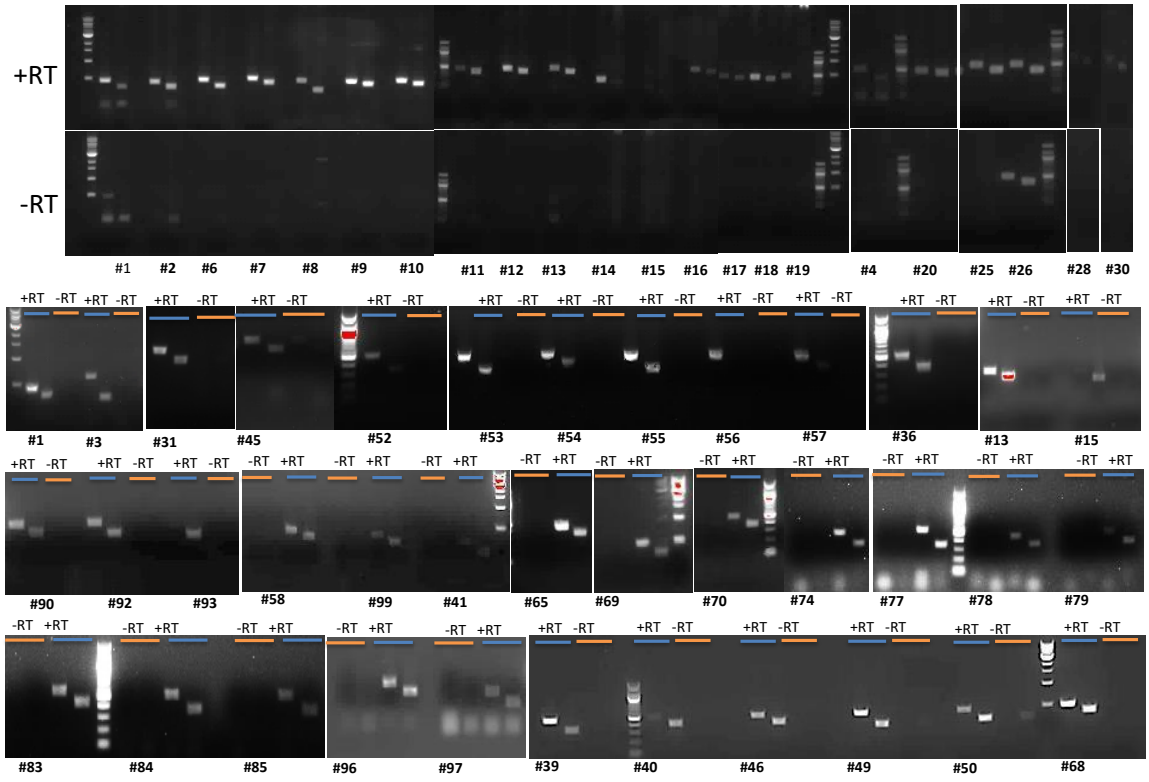


Figure 2.4: Experimental analysis of 5′UTRs by RT-PCR.

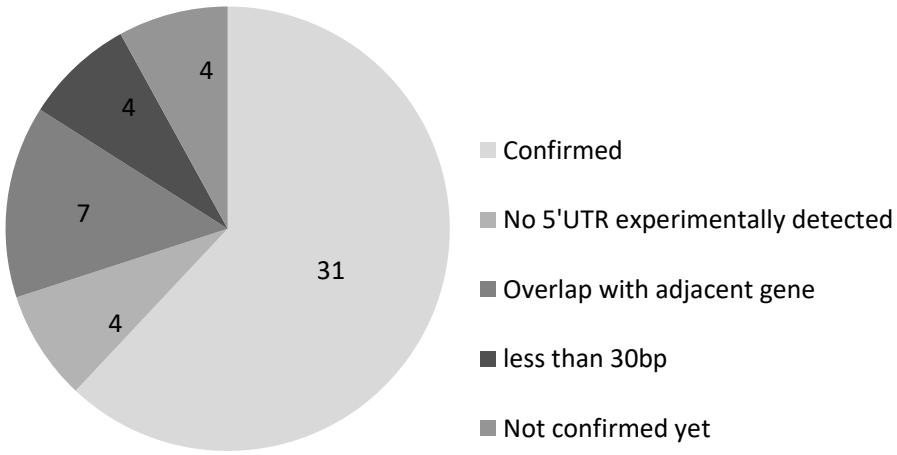


Figure 2.5: Summary of the results of 5′ RACE.

2.3.4 Establishment and Validation of the High-Throughput Fluorescence-Based Screening System

To test each 5'UTR candidate's ability to regulate downstream gene expression, an *in vivo* fluorescence-based reporter gene screening system was developed, taking advantage of previous successful efficient expression of GFP in *Z. mobilis* [86]). As shown in **Figure 2.6A**, inducible expression of GFP was confirmed under the P_{tet} promoter, and then the 60-nucleotide sequence of the theophylline riboswitch element was cloned in front of the GFP gene to establish the functionality of the GFP screen in *Z. mobilis* (**Figure 2.6B**). **T**he fluorescence shifted when it is induced with 10 μg/mL tetracycline compared to that of the un-induced control sample (**Figure 2.6C**). After confirming the functionality of the fluorescence system in *Z. mobilis*, the well-characterized theophylline synthetic riboswitch was used as a test case to establish the screening system that was able to elicit a fluorescence change specific to the activation of a 5'UTR. The theophylline riboswitch has been engineered as a synthetic riboswitch system to control gene expression in various bacterial species and natively controls gene expression at the translational level by binding to the small molecule, theophylline [87–89].

To establish the functionality of the GFP screen, the 60-nucleotide sequence of the theophylline riboswitch element was cloned in front of the *gfp* gene in *Z. mobilis* (**Figure 2.6B**) and the level of GFP expression with 2 mM theophylline was compared with a DMSO control in *Z. mobilis*. Control-GFP exhibited about 10-fold GFP expression when it was induced (**Figure 2.6C,D**). In contrast to control-GFP, a 2-fold increase in GFP expression from theophylline RNA stability element (RSE)-GFP construct was observed. Although the fold change in GFP was not high (limited to about 2-fold), this was enough to screen for positively activating candidates. It is also worth noting that this is not an unusual fold change of expression when using the same theophylline riboswitch

in different bacteria species [89] without engineering pairing strength of the region between the aptamer and RBS (Ribosome Binding Site) in the theophylline switch within a particular species. Importantly, the activation of the theophylline switch was successfully demonstrated, as well as the use of this 5'UTR *in vivo* fluorescence-based screening system in *Z. mobilis* for the first time. This could be a useful tool for screening the control of gene expression in metabolic pathways related with ethanol tolerance or other stress responses in *Z. mobilis*.

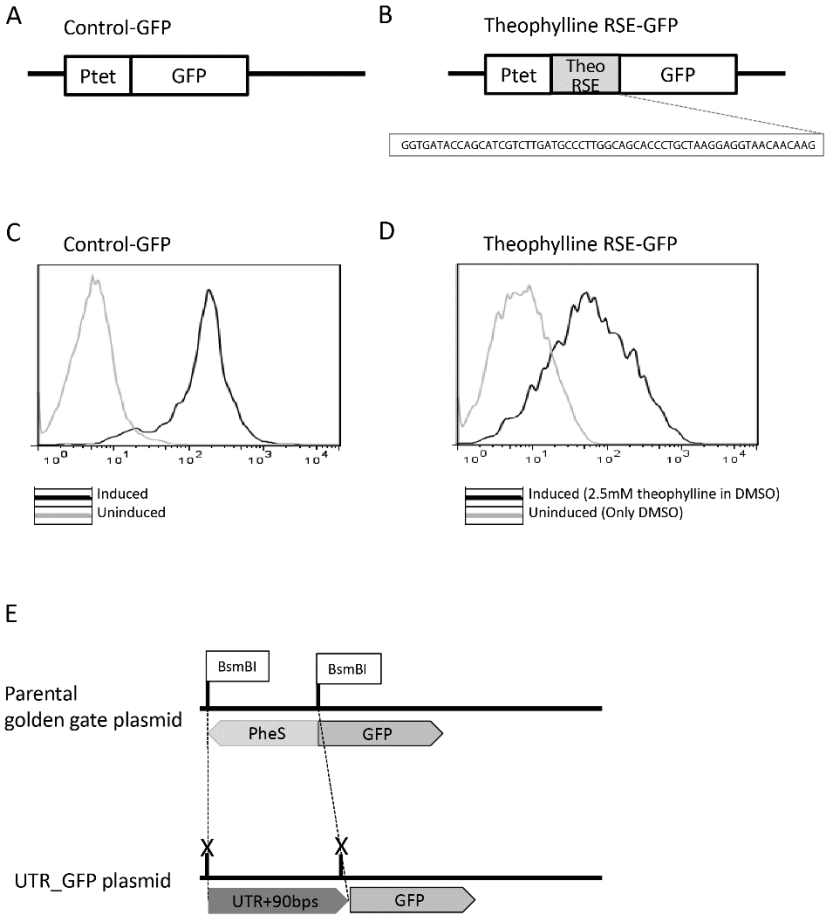


Figure 2.6: Establishment of a GFP-based high-throughput reporter gene screening system to characterize regulatory 5′UTR regions.

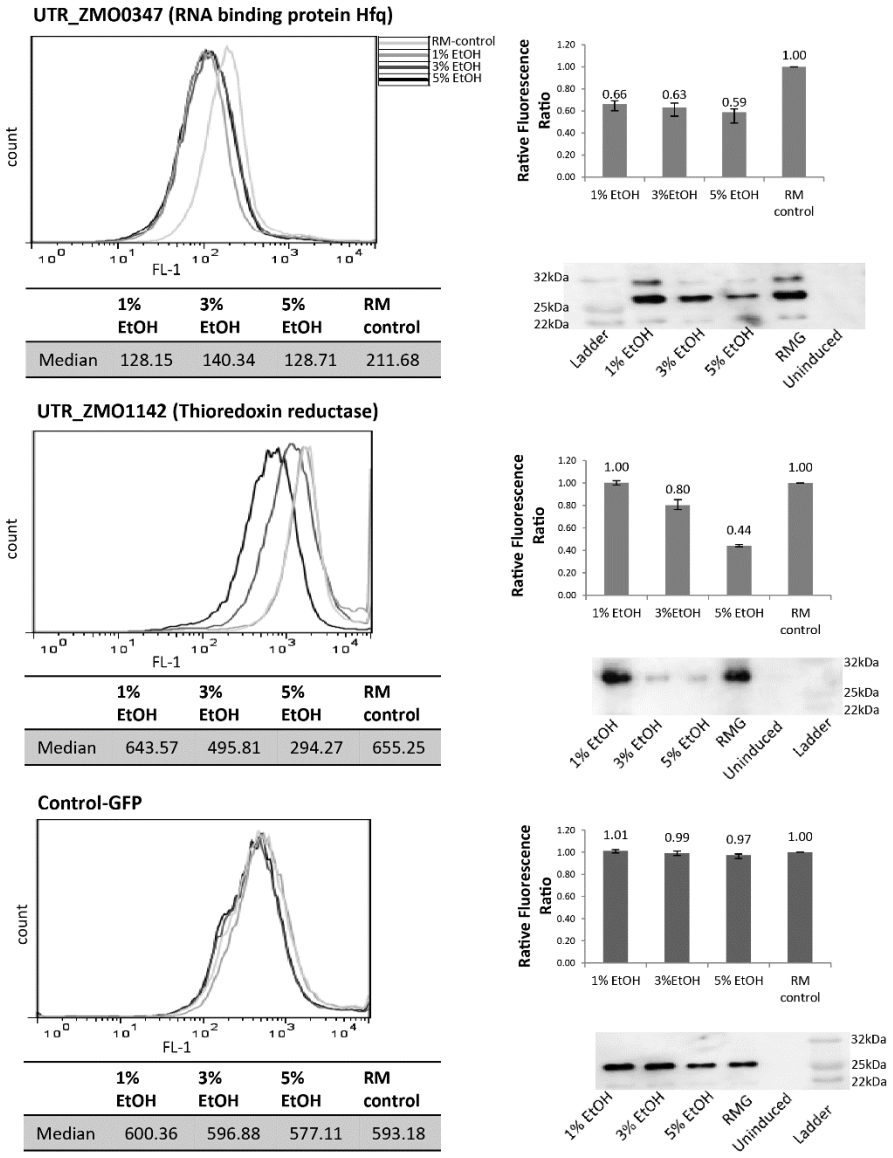
Control-GFP plasmid was developed in which GFP is located under P_{tet} promoter control to demonstrate expression of the GFP protein in *Z. mobilis* **(A)**. The theophylline synthetic riboswitch was cloned in front of GFP to regulate its expression and show that a well-characterized regulatory 5′UTR could be used to see measurable differences in GFP expression upon the addition of the ligand it senses **(B)**. Upon induction of P_{tet} with tetracycline (black), the Control-GFP strain shows a fluorescence shift by flow cytometry compared to the un-induced sample (light gray) **(C)**. The strain containing plasmid theophylline RNA stability element (RSE)-GFP was induced with 2.5 mM theophylline in DMSO (black) and compared by flow cytometry to a sample to which only DMSO (light gray) was added. In both cases, P_{tet} was induced with tetracycline **(D)**. To allow screening of 5′UTRs identified in this study, the plasmid was designed with BsmBI sites flanking the PheS cassette in front of GFP. The PheS cassette was replaced with each UTR+90 bps of initial mRNA sequence by Golden Gate cloning **(E)**.

2.3.4 Identification and Characterization of Stress-Responsive Regulatory 5'UTRs Using Reporter Gene System

Utilizing the *Z. mobilis* pEZ minimal shuttle vector system [90], a high-throughput cloning strategy was developed for construction of each 5'UTR-containing GFP plasmid by the combination of Golden Gate assembly and PheS counter selection marker (**Figure 2.6E**). Given that nucleotides in the coding region may affect the structure of 5'UTRs for the regulation of the gene, each 5'UTR sequence (verified by 5' RACE) was cloned with an additional 90 base pairs of the downstream coding region of the corresponding gene for the generation of 5'UTR-GFP libraries. Primers used for the generation of 5'UTR-GFP libraries are listed in **Table A.3**. Using this plasmid construct, a library of 36 5'UTR-GFP candidates was generated (**Table 2.1**) for further testing. The sequences and features of the final 36 candidates were shown in **Table A.2**.

Since high tolerance to ethanol is one of the desirable features of *Z. mobilis*, the initial screen of 5'UTR-GFP libraries was conducted under ethanol stress (5% (v/v) ethanol-supplemented media) to evaluate potential 5'UTR activation relative to a standard medium (RM) control. A strain with only GFP and no 5'UTR (Control-GFP) was used as a negative control. All experiments were done in biological triplicates. After the signal difference was observed to be highest at 10 h post-induction, the fluorescence measurements were taken at this time point for all other experiments. This is consistent with previous data that showed maximum GFP fluorescence in late exponential phase [86]. Upon screening all candidates, two were identified that exhibited significant fluorescence changes under 5% ethanol supplementation compared to the control without added ethanol: the 5'UTR of ZMO0347 (RNA binding protein Hfq, UTR_ZMO0347) and the 5'UTR of ZMO1142 (thioredoxin reductase, UTR_ZMO1142) (**Figure B.1**).

To explore responsiveness of these 5'UTRs under different levels of ethanol, 1, 3, or 5% (v/v) of ethanol were added to the cultures before fluorescence and western blotting analysis (**Figure 2.6**). In this plasmid system, GFP fluorescence uniformly decreased under any level of ethanol stress when regulated by the UTR_ZMO0347 (RNA binding protein Hfq), whereas the fluorescence under the control of UTR_ZMO1142 (thioredoxin reductase) showed an ethanol concentration-dependent decrease (**Figure 2.6**). The changes in GFP expression in cells were also confirmed by western blotting analysis, which corresponded well with fluorescence data (**Figure 2.6**). All 5'UTR candidates were also tested for responses to xylose and acetate stresses in this fluorescence screening system (**Figure 2.8**). Interestingly, UTR_ZMO1142 did not appear to regulate gene expression under xylose and acetate stresses, but only under ethanol stress (**Figure 2.8**). Consistent with these results, it has been reported that thioredoxin reductase (ZMO1142) is less abundant under 6% ethanol stress at the protein level, although not at the transcript level, indicating a layer of translational regulation [52]. Additionally, for Hfq (ZMO0347), a protein that is highly associated with stress response in various organisms, a decrease of transcripts has been previously detected under 6% ethanol stress [52,91,92]. However, inconclusive data about changes in Hfq protein levels are found throughout the literature (and our own studies) given the difficulty in detecting this protein by mass spectrometry. Importantly, the responses of UTR_ZMO0347 and UTR_ZMO1142 to ethanol indicate that UTR regulation could be part of a mechanism of ethanol stress protection in *Z. mobilis*.



	1% EtOH	3% EtOH	5% EtOH	RM control
Median	128.15	140.34	128.71	211.68

	1% EtOH	3% EtOH	5% EtOH	RM control
Median	643.57	495.81	294.27	655.25

	1% EtOH	3% EtOH	5% EtOH	RM control
Median	600.36	596.88	577.11	593.18

Figure 2.7: Identification of 5′UTRs that regulate GFP expression under ethanol stress.

UTR_ZMO0347 and UTR_ZMO1142 showed fluorescence shifts under different levels of ethanol compared to control-GFP. White gray line: RMG (normal media). Light gray line: 1% ethanol supplemented media. Dark gray line: 3% ethanol supplemented media. Black line: 5% ethanol supplemented media. Median fluorescence level of each condition is shown in a table and the fluorescence ratios relative to the RMG control are shown in the bar graphs. GFP protein levels were detected by western blotting using anti-GFP antibody. These experiments were done in triplicate and the fluorescence ratio bar graph error bars represent standard deviation from the mean.

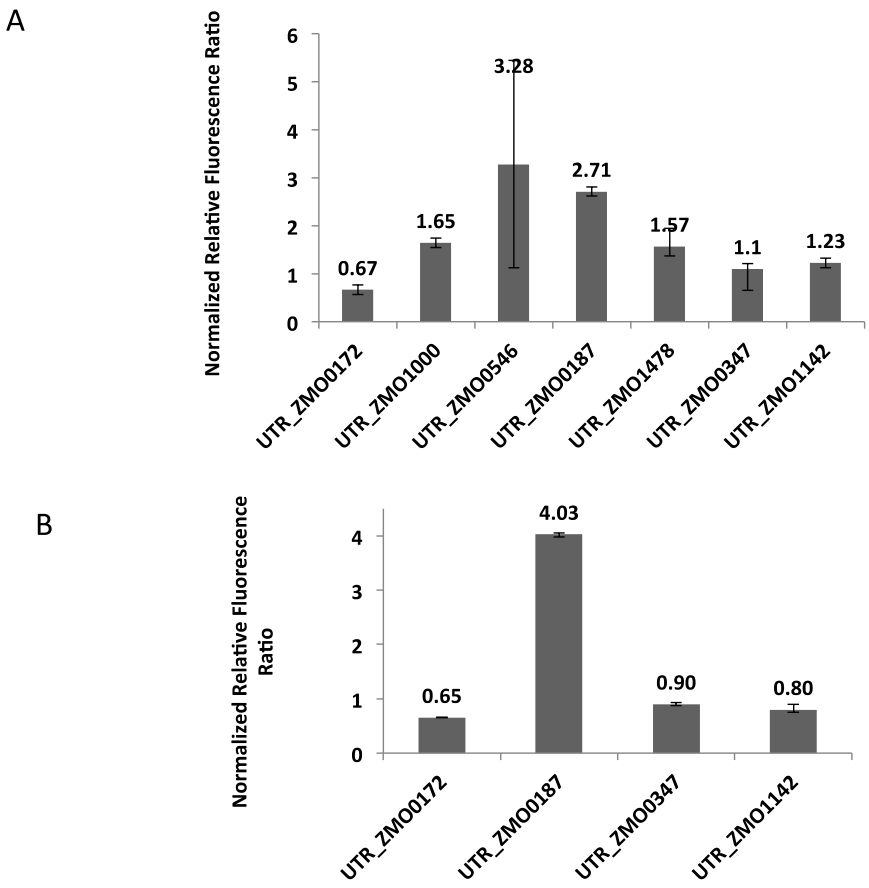


Figure 2.8: The effect of acetate (A) and xylose (B) on 5′UTRs.

In addition to ethanol stress responsive 5'UTRs, a few acetate and xylose responsive elements were identified (**Figure 2.8**). A previous study confirmed that regulatory mechanisms responding to acetate stress and xylose utilization mainly included differential expression of carbon and energy metabolism genes to reduce the impact of stress on the cell [64]. Acetate stress repressed genes associated with flagellar system and glycolysis, but induced genes related to stress responses and energy metabolism [51]. 6-phsphogluconolactonase catalyzes hydrolysis of the ester linkage of lactone, resulting in production of 6-phophogluconate in the pentose phosphate pathway. Up-regulation of

ZMO1478 (6-phosphogluconolactonase) transcript in acetate stress has been previously reported [64], and in this study, UTR_ZMO1478 (6-phosphogluconolactonase) was shown to activate GFP expression under acetate stress (10 g/L sodium acetate). Taken together, these results suggest that 5'UTR could be responsible for the natural increase of ZMO1478 expression under acetate stress. Additionally, UTR_ZMO0172 (thiamine biosynthesis protein) showed down-regulation of GFP expression under both acetate and xylose stresses, indicating that this gene could be involved in general stress response mechanisms. Further studies on protein expression levels of ZMO0172 under stressful conditions could help reveal the underlying regulatory mechanism of UTR_ZMO0172.

2.3.5 Native Regulation of Hfq Translation by 5'UTR

To evaluate the physiological effects of ethanol stress-responsive regulatory 5'UTRs in *Z. mobilis*, a mutant strain was constructed with a genomic deletion of UTR_ZMO0347. Unfortunately, UTR_ZMO1142 could not be deleted due to the limitation of the technique used in this study for short length deletion (e.g., 57 bp in this study for UTR_ZMO1142). The UTR_ZMO0347 deletion was constructed utilizing a homologous recombination deletion technique to disrupt the 5'UTR region (except for promoter and RBS regions) of the *hfq* gene in *Z. mobilis* 8b strain. The spectinomycin resistance gene was used as a selection marker between 1 kb up- and down-stream homology arms. The deletion of this region was confirmed by PCR and Sanger sequencing (data not shown). Primers used for the construction and confirmation of deletion constructs are shown in **Table A.3**. Growth data for the *Z. mobilis* parental strain 8b and its *hfq* 5'UTR deletion mutant strain were collected (**Figure 2.9**). A growth defect in the ΔUTR_ZMO0347 strain was observed relative to the wild type strain under no ethanol stress.

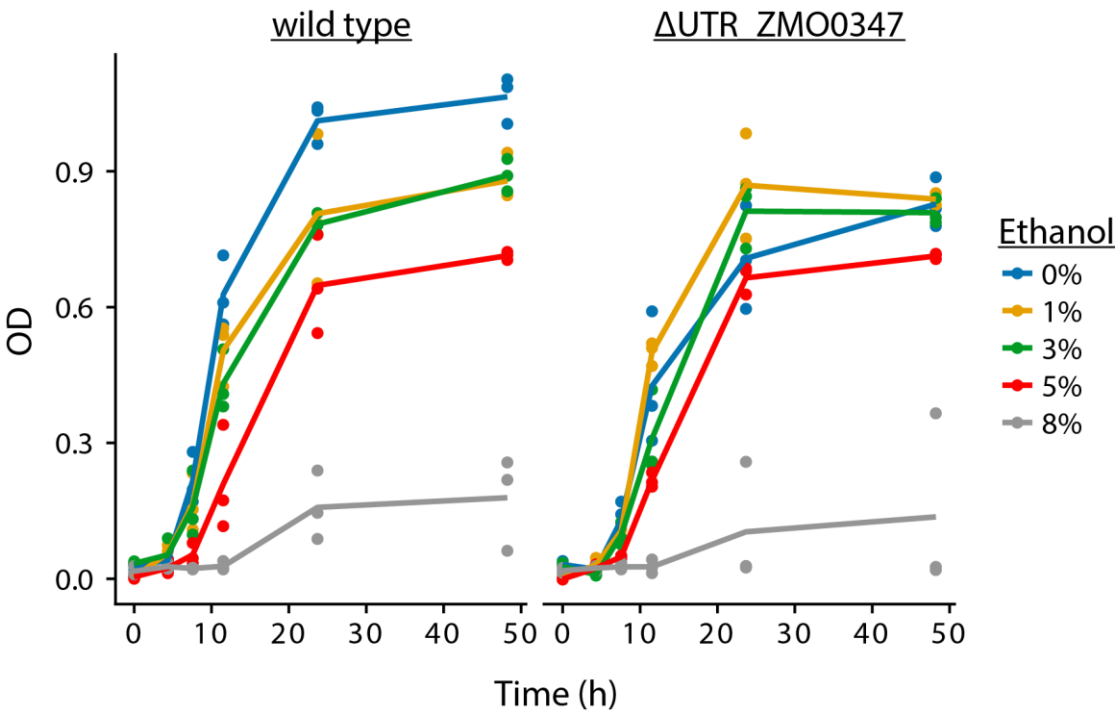


Figure 2.9: Growth curves of the wild type and ΔUTR_ZMO0347 strains under a range of ethanol stresses.

Transcript levels of the ZMO0347 (Hfq) mRNA in the deletion strain were measured via qRT-PCR with and without 5% (v/v) ethanol supplementation (**Figure 2.10A**). Although the deletion strain shows an overall decrease in transcription, a consistent up-regulation of the *hfq* transcript is observed in both the wild type and deletion mutant strains in the presence of ethanol. This suggests that the ethanol-responsive regulatory effect of the 5'UTR is not at the transcriptional level (e.g., due to promoter effects).

At the protein level, native Hfq levels were also evaluated in the ΔUTR_ZMO0347 mutant strain relative to its wild-type ZM4 strain via western blotting analysis (in chromosomally FLAG-tagged Hfq strains for both the wild type ZM4 and the

ΔUTR_ZMO0347 strain). As shown in **Figure 2.10B,C** for the wild type and deletion strains respectively, the Hfq protein showed a small (less than a 2-fold increase) change in response to ethanol. In the wild type strain, increases in Hfq protein expression were observed starting with 5% (v/v) ethanol supplementation (but not with 1% ethanol supplementation). In contrast, in the ΔUTR_ZMO0347 deletion strain, a more sensitive response was observed upon supplementation with ethanol, in which the increase in Hfq protein was observed with only 1% (v/v) ethanol supplementation (**Figure 2.10**). This suggests that the native 5'UTR region serves as a post-transcriptional regulator, akin to a resistor, that keeps Hfq protein levels down-regulated in response to lower ethanol stress. Given the fact that the isolation of 5'UTRs in our GFP assays (in the absence of their native promoters and genetic context) indicated down-regulation of protein expression by the 5'UTR, the overall increase of *hfq* transcript and protein levels in the wild type strain with ethanol suggests that additional regulatory effects play a role in overall Hfq regulation. As discussed below, this is not surprising given that, in *E. coli* and other organisms, Hfq is globally regulated by highly complex schemes that involve its native promoter and global protein-sRNA regulators [93–96].

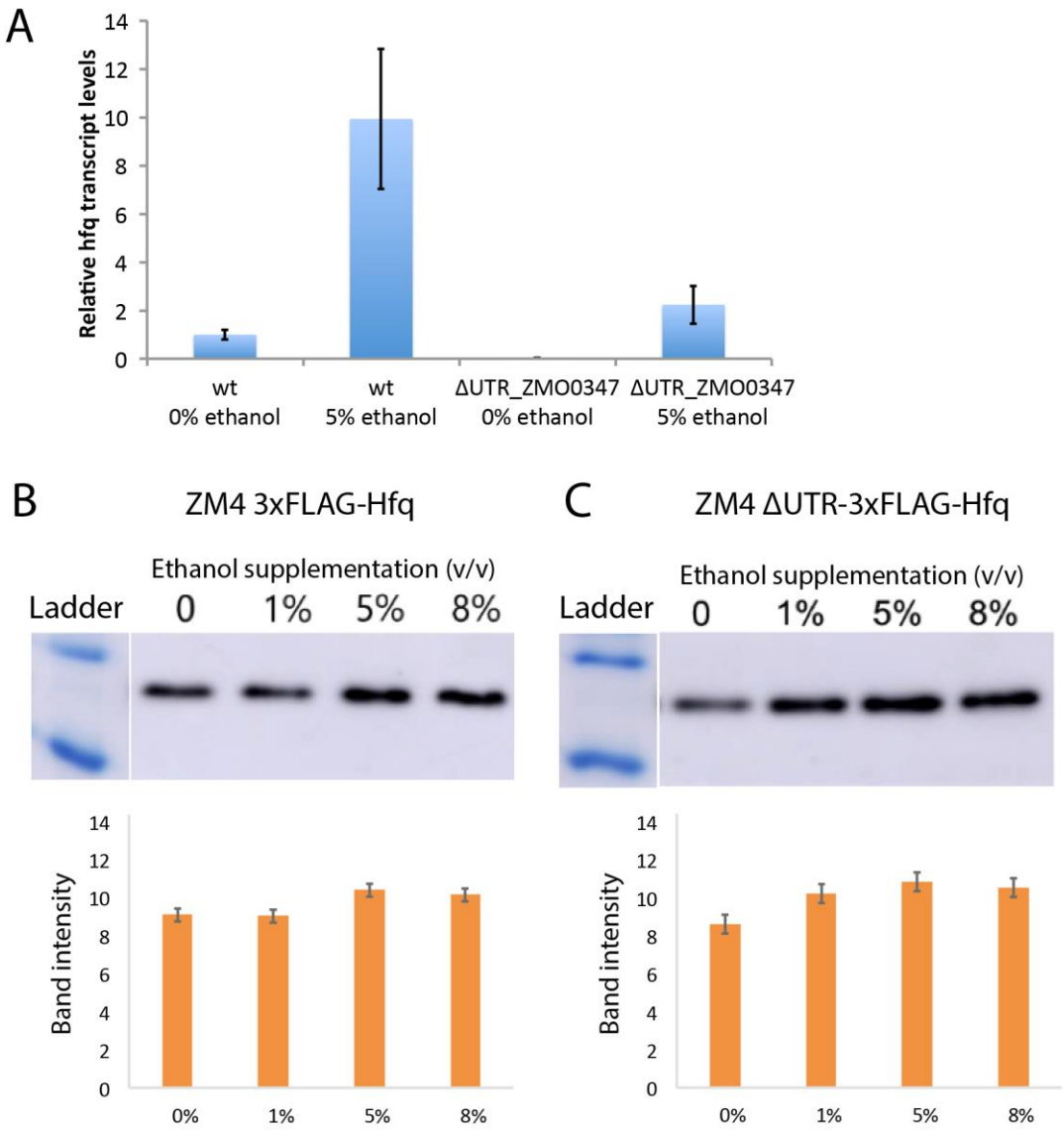


Figure 2.10: Native regulation of Hfq in ethanol stress.

Transcript levels of *hfq* were quantified by qPCR with (5%) and without (0%) ethanol supplemented to the media **(A)**. Bars represent averages of biological triplicates calculated by the comparative delta-delta threshold cycle ($\Delta\Delta C_T$) method and error is standard deviation. The Hfq protein levels were quantified by western blots of chromosomally FLAG-tagged Hfq strains with the UTR **(B)** and without **(C)**. After growth in a range of ethanol concentrations, 20 ng of total protein were loaded into an SDS-PAGE gel, then transferred to a membrane, and blotted with the anti-FLAG antibody. Band intensity was quantified and is shown as averages of biological triplicates with error as standard deviation.

2.4 DISCUSSION

Bacterial 5'UTR elements contribute to comprehensive gene regulation under stress conditions. They rapidly sense and respond to environmental changes in order to orchestrate a cascade of gene expression and protein activity changes [97]. Conventionally, sequence-based conservation analysis such as Rfam has been widely used to identify UTRs. However, since it is based on alignments of UTRs across organisms, this kind of approach is limited to known UTRs and to the identification of functional homologs from closely related species. Therefore, as in this study, using high-throughput transcriptomic data to identify novel 5'UTRs and functional homologs of known 5'UTRs in less-studied and non-model bacteria could significantly improve the identification of 5'UTRs and to screen the functioning of these 5'UTRs under a variety of environmental (e.g., temperature and pH) or intracellular conditions (i.e., changes in particular small molecule concentration). This work thus highlights the feasibility of using transcriptomic data, bioinformatics analysis, and fluorescence-based screening to identify novel regulatory 5'UTRs in other microorganisms.

Additionally, this study has demonstrated that 5'UTR elements in *Z. mobilis* have regulatory roles, particularly under ethanol stress conditions. UTR_ZMO0347 (RNA binding protein Hfq) and UTR_ZMO1142 (thioredoxin reductase), which are associated with ethanol stress responses, showed down-regulation of the downstream genes. Interestingly, this regulatory effect was observed more drastically under higher concentrations of supplemental ethanol for UTR_ZMO1142. The downstream gene of UTR_ZMO0347 encodes a homolog of RNA binding protein Hfq, an RNA chaperone and regulator of the small RNA network. Hfq also mediates transcription anti-termination via binding to Rho factor for the control of gene expression at the transcriptional level in *E. coli* [98]. Previous studies also indicated that this Hfq homolog in *Z. mobilis* is associated

with stress responses [48]. The data in this study thus suggest that UTR_ZMO0347 mediates regulatory effects on its downstream gene *hfq*, even though it is not clear that this effect is induced by direct interaction with ethanol or through the participation of other factors involved in the stress response. Thioredoxin reductase (ZMO1142) catalyzes the reduction of thioredoxin coupled with NADPH and as such plays a major role in the defense mechanism for the oxidative stress via the reduction of disulfide bonds by thioredoxin reductase [99]. UTR_ZMO1142 also responded to xylose and acetate stresses, implicating it as a potential regulator of the general stress response.

Because of the response of UTR_ZMO0347 (Hfq) to ethanol stress, this UTR was genetically deleted. Two major phenotypic differences were observed between the UTR mutant and the wild type strains: (1) a general decrease in *hfq* expression at the transcript level, and (2) up-regulation of Hfq protein expression within a wider range of ethanol stress levels specifically at lower levels (**Figure 2.11**). In addition to the translational regulation, this work also demonstrates that the presence of the 5'UTR affects transcriptional levels of *hfq*, likely by providing stability to its transcript. In *Staphylococcus aureus*, Hfq protein was identified as a global regulator involved in stress and pathogenicity [100], and it is suspected that this protein could have a similar role throughout bacteria. As an important regulatory factor upon stress, it has been reported that Hfq was also involved in osmotic and ethanol stress in *L. monocytogenes* [101], where the transcription of *hfq* was induced upon ethanol and osmotic stress depending on sigma factor σ^B for the regulation of stress responses in *L. monocytogenes* [101]. However, regulation of *hfq* itself and its UTR still needs to be elucidated, particularly in response to various stresses in different growth phases in *Z. mobilis*.

Interestingly, in this work, the opposite trend of the effect on the native 5'UTR on expression of the *hfq* transcript and (mildly) on the Hfq protein was observed (increase

expression in presence of 5'UTR) relative to the effect observed at the protein level with our heterologous reporter, where GFP consistently decreased when regulated by the isolated 5'UTR. This observation can be attributed to the complexity of regulatory mechanisms known to affect *hfq*. A potential model of the contribution of the 5'UTR for native Hfq regulation is illustrated in **Figure 2.11**. With the 5'UTR intact, transcript levels and the protein produced are higher with 5% ethanol supplementation than that without external addition of ethanol. Without the 5'UTR, the transcripts levels are lower, but show the same trend as with the UTR intact. Interestingly, the protein is upregulated at a lower ethanol concentration (1%) when the UTR is missing compared to when it is present (upregulation not seen at 1%, but at 5%) (**Figure 2.10**). This suggests that the Hfq protein levels are more sensitive to upregulation without the 5'UTR. The UTR's post-transcriptional regulation role may be to minimize over-production of Hfq protein from the more abundant transcripts in mild ethanol stress. Additionally, although the transcripts are fewer in the mutant strain with the Hfq 5'UTR deleted, protein production is about the same in both strains at 5% ethanol, suggesting post-transcriptional regulation by the 5'UTR.

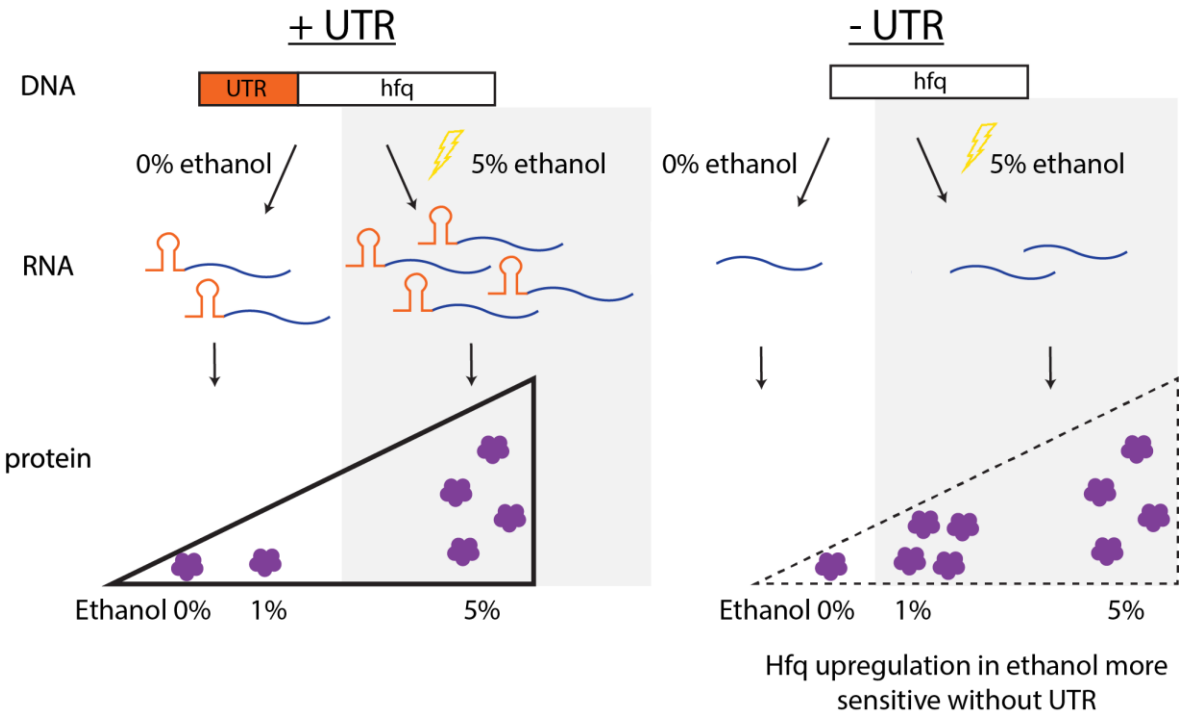


Figure 2.11: Summary of observed Hfq regulation in ethanol stress.

In native conditions with the UTR of Hfq intact (ZMO0347), the transcript and protein levels increase in 5% ethanol stress. When the UTR is deleted from the genome, the protein expression increases upon a more mild ethanol stress (1%), indicating the sensitively of Hfq expression without the UTR.

Overall, this study provided a strategy to identify regulatory 5'UTR using omics datasets and to characterize them using a fluorescence reporter gene system for bacterial species. It also shed a light on the regulatory role of 5'UTR regions in stress responses in *Z. mobilis* for the first time, after constructing and validating a novel genome-wide 5'UTR screening tool in this organism. A regulatory 5'UTR region on *hfq* has been identified and the regulation of Hfq in particular at the protein level appeared to be important in viability under ethanol stress for *Z. mobilis*, even though complex regulatory networks around the *hfq* gene should be further investigated. This work also provided novel responsive

regulatory 5'UTR biological parts for future metabolic engineering efforts to enhance microbial robustness and temporal gene expression.

Chapter 3: Small Noncoding RNAs Involved in Ethanol Tolerance in *Zymomonas mobilis*

3.1 Chapter Summary

sRNAs represent a powerful class of regulators that influence multiple mRNA targets but remain largely uncharacterized outside of model organisms. *Zymomonas mobilis* is a natural ethanol-producing bacterium in which multiple small RNAs (sRNAs) have recently been identified, some of which show differential expression in ethanol stress. In this study, we show that sRNAs Zms4 and Zms6 have significant impacts on ethanol tolerance in *Z. mobilis*. We have conducted multi-omics analyses (including proteomics, transcriptomics, and sRNA-immunoprecipitation) to map gene networks under the influence of their regulation. Zms4 and Zms6 represent the first sRNAs with regulatory functions confirmed in *Z. mobilis* and are shown here to bind to multiple mRNA targets. Specific binding sites of two mRNA targets have been defined. Interestingly, Zms4 and Zms6 interact with each other as well as two other sRNAs, forming a novel sRNA-sRNA interaction network. To date, very few direct sRNA-sRNA interactions have been reported in any organism. Our results indicate that this uncovered sRNA network co-orchestrates multiple ethanol tolerance pathways through a diverse set of mRNA targets including transport, oxidoreductase activity, and lipid metabolism pathways.

3.2 Introduction

As global controllers of gene expression, small RNAs (sRNAs) represent powerful tools for engineering complex phenotypes [17,24,96]. These typically non-coding RNAs are 50-300 nucleotide (nt) transcripts that act as regulators of mRNA and protein expression, mostly by blocking translation or changing stability of mRNAs [12]. Although less common, proteins have also been shown to be targets of sRNA regulation [102]. Traditionally thought

of as non-coding RNAs, many have been discovered in intergenic regions [103], although some are now known to produce small peptides [104]. The majority of known sRNAs act in *trans*, meaning their targets are encoded elsewhere in the genome, whereas *cis*-encoded sRNAs neighbor their targets in the genome on the same or opposite DNA strand [105].

Advances in high-throughput sequencing have enabled discovery of hundreds of sRNAs across bacteria [106,107], but characterization lags far behind. As a result, the vast majority of sRNAs have functions completely unknown, especially in non-model organisms. Current approaches to determine sRNA target networks take advantage of RNA-seq and proteomics [108,109], although a challenge remains to decouple direct vs indirect interactions. Computational tools such as IntaRNA can be helpful in predicting the most favorable sRNA-mRNA interactions and their binding sites, although in vivo conditions and competition of multiple targets for binding sites cannot be accounted for [110]. Ultimately, electrophoretic mobility shift assays (EMSAs) have been useful in testing direct binding of RNA and protein targets in vitro.

Increasingly complex regulatory networks have been discovered, including a few direct sRNA-sRNA interactions. One reported interaction is between sRNAs RyeA and RyeB in *E. coli*, which originates from the same intergenic region, encoded in opposite directions [111]. However, due to the complete overlap of RyeB in antisense sequence to RyeA, the binding of these sRNAs is expected and it results in RNase III-dependent cleaving [111]. Although still largely uncharacterized in *Escherichia coli,* the target network of RyeB's homolog SdsR has been characterized in *Salmonella enterica* and includes stress response regulators [112]. Another known interaction is between sRNA GcvB and the RNA sponge SroC, which represses GcvB [113]. This mRNA cross-talk forms a feed-forward loop in the regulation of ABC transporters and affects growth in different nutrient conditions [113].

An advantage of sRNA regulation is its efficiency compared to protein regulators like transcription factors because they do not require translation and act directly on mRNA transcripts [7]. Their dynamic nature and low metabolic burden make sRNAs especially suitable to coordinate stress responses including temperature, nutrient, membrane, oxidative, iron, pH, and anaerobic stresses [4,5]. Ethanol tolerance represents a complex phenotype that sRNAs appear to help regulate as well. For example, sRNA Nc117 in *Synechocystis* sp. PCC 6803 [114] as well as OLE RNA in *Bacillus halodurans* C-125 [76] both appear to protect the cells from ethanol toxicity. However, the mRNA and/or protein targets of these sRNAs are unknown (Nc117) or limited in number (OLE RNA). OLE RNA is known to bind to RNase P as well as a protein (aptly named the OLE-associating protein), which associates to the membrane [75,76,115]. The lack of network characterization in these contexts has precluded advances in understanding alcohol tolerance and in general sRNA function in non-model organisms. Moreover, as it relates to the specific phenotype of ethanol tolerance, these uncharacterized ethanol-related regulatory RNAs have left unanswered the questions of the specific pathways that are co-regulated to naturally grant the ethanol resistance phenotype in some organisms.

Zymomonas mobilis is an especially biotechnologically-relevant bacterium due to its natural ethanol producing ability up to 12% (w/v) and ethanol tolerance up to 16% (w/v) [34,35]. Over the last 20 years, a variety of *Z. mobilis* strains have been developed through metabolic engineering and directed evolution [59,60]. The responses of *Z. mobilis* to a variety of stresses including ethanol, furfural, and acetate have been explored by transcriptomics and proteomics approaches [48,50–52,63]. Each of these stress responses is considered a complex phenotype because they trigger the differential expression of large sets of transcripts and proteins with a wide variety of cellular functions. For example, the ethanol stress response has been characterized to include up regulation of protein folding

chaperones, DNA repair proteins, and transporters and down regulation of genes related to translation, ribosome biogenesis, and metabolism [52]. However, regulation mechanisms that contribute to these widespread changes remain unclear. It is likely that these complex phenotypes are made possible by multiple layers of regulation (DNA, RNA, protein) coordinating responses to extracellular environments. Recently, 15 sRNAs were discovered in *Z. mobilis*, of which 4 were shown to have differential expression to anaerobic or ethanol stresses, representing ideal systems to further investigate another specific and distinct sRNA-driven regulatory network in a non-model organism [53].

In this study, we show that two sRNAs, Zms4 and Zms6, have significant impacts on ethanol tolerance and production in *Z. mobilis*. Multi-omics analyses were used to map their regulatory networks including mRNA, protein, and other sRNA targets. Our results indicate that this uncovered sRNA network co-orchestrates multiple ethanol tolerance pathways through a diverse set of mRNA targets including transport, oxidoreductase activity, and lipid metabolism pathways. This work presents the first sRNAs with regulatory binding interactions confirmed in *Z. mobilis*, as well as the largest known sRNA-sRNA interacting network in bacteria.

3.3 RESULTS

3.3.1 sRNAs Affect Ethanol Tolerance and Production

Four sRNAs (Zms2, Zms4, Zms6, and Zms18) were previously shown as differentially expressed under 5% (v/v) ethanol stress and/or anaerobic conditions [53], favoring glucose consumption, enabling higher growth rates and ethanol accumulation [48]. To determine their direct impacts on the ethanol tolerance phenotype, we overexpressed and deleted each sRNA. However, Zms2 could not be independently overexpressed because it significantly overlaps with its neighboring gene ZMO1198 and Zms18 could not

be fully deleted due to multiple homologous regions in the genome. As such, we focus on characterization of Zms4 and Zms6 in this study. After validation of each sRNA overexpression and deletion strain (**Figures 3.1 and 3.2**), these strains were grown anaerobically in media with and without 8% (v/v) ethanol supplementation in 300 µL Bioscreen C cultures to observe the effects of each sRNA on the phenotype. This concentration of 8% (v/v) ethanol was selected to significantly impact cell growth while still maintaining viability (and therefore reproducibility); *Z. mobilis* is expected to have ~30% of the normal cell density at saturation under these conditions [52]. As shown in **Figure 3.3A**, the overexpression of Zms6 significantly increases growth rate in ethanol stress compared to the empty plasmid and wild type controls, while the growth rates in the unstressed condition are relatively unchanged. Additionally, ΔZms4 and ΔZms6 strains have significantly reduced growth under ethanol stress. Interestingly, previous work showed Zms6 to naturally increase in expression levels when 5% (v/v) ethanol was added to the media [53], indicating the possibility that the natural regulation of its expression plays a role in increasing robustness to ethanol stress.

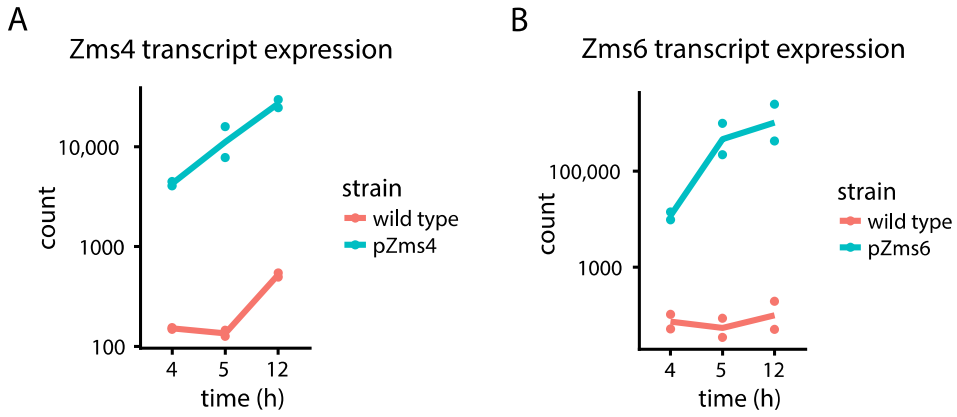


Figure 3.1: Verification of transcript levels in sRNA overexpression strains.

Transcript counts (normalized by DESeq2) show inducible overexpression of Zms4 (A) and Zms6 (B) over time. Each point represents a biological duplicate and the line their average.

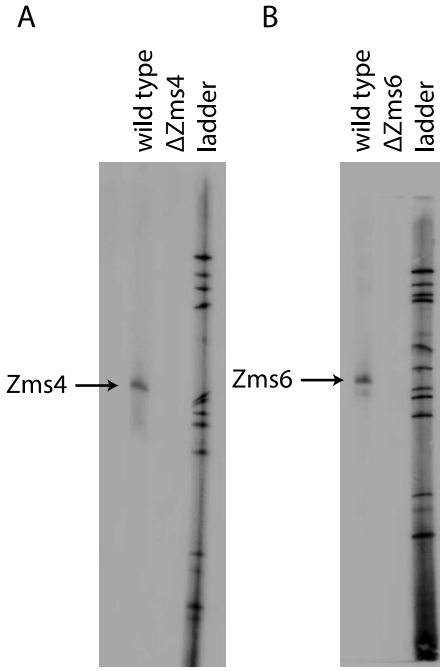


Figure 3.2: Verification of transcript levels in sRNA deletion strains.

Northern blots show successful deletion of Zms4 (A) and Zms6 (B).

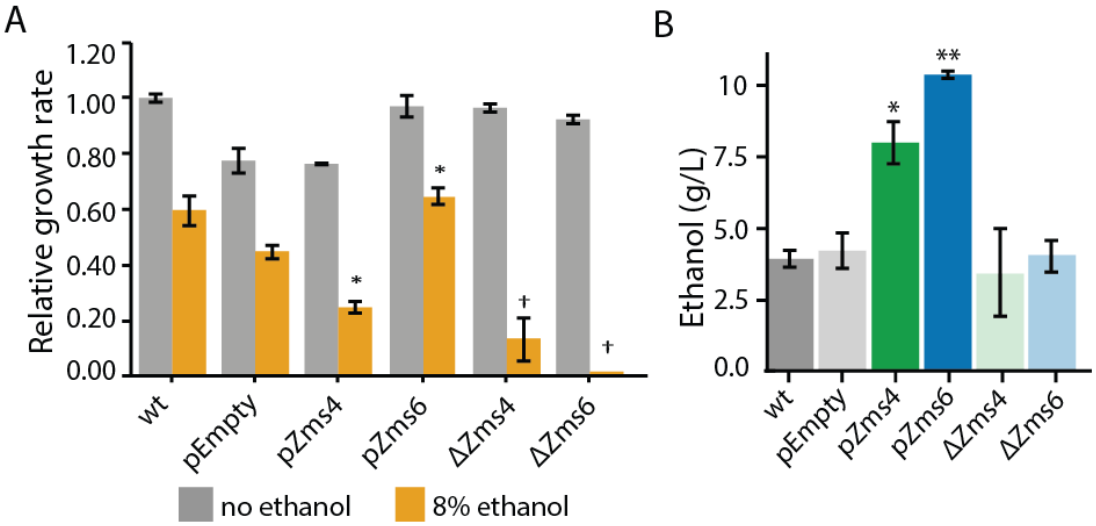


Figure 3.3: sRNA expression levels affect ethanol tolerance and production.

(A) Mutant strains overexpressing (pZms) and lacking (ΔZms) Zms4 and Zms6 were grown with and without 8% (v/v) ethanol supplementation grown in 300 µL cultures in the Bioscreen C. Values represent mean ± SD of biological triplicates, normalized to the wild type growth rate in the no ethanol condition. Significant values are *P ≤ 0.001 from Student's t-test of each strain compared to pEmpty (for overexpression strains) or †P ≤ 0.001 wild type (for deletion strains). (B) Concentration of extracellular ethanol of 500 mL cultures at 8 h (exponential phase). Cells were grown anaerobically at 33°C and sRNA expression from the plasmids was induced at 4.5 h. Values represent mean ± SD of biological duplicates. Significant values are *P ≤ 0.05 or **P ≤ 0.01 from Student's t-test of each strain compared to pEmpty.

Considering the natural ethanol production of *Z. mobilis* and the tolerance effects observed in **Figure 3.3A**, we reasoned that the regulatory contribution of Zms4 and Zms6 to the high ethanol tolerance phenotype could also be captured by measuring ethanol production. As shown in **Figure 3.3B**, Zms4 and Zms6 overexpression strains show significantly higher levels of ethanol production relative to the empty plasmid and wild type strain.

3.3.2 Multi-omics Analyses Reveal Gene Networks Associated with Zms4 and Zms6

As global regulators, sRNAs interact with multiple mRNA and protein targets to coordinate complex phenotypes. Given that sRNAs can repress or activate gene expression, we expect that altering their stoichiometry will lead to significant changes on the innate levels of their direct or indirect targets. To uncover the networks of targets affected by Zms4 and Zms6, we used an Integrative FourD-Omics (INFO) approach [96], useful for identification of regulated targets and regulatory mechanisms underlying sRNA-driven systems. This integrated analysis (**Figure 3.4**) included collection of transcriptome, proteome and MS2-affinity purification coupled with RNA sequencing (MAPS) profiles evaluated in Zms4 and Zms6 overexpression strains, inducible by a P_{tet} promoter. Cells were collected before induction during exponential-phase (4 h), 1 hour post-induction (5 h), and 7 hours post-induction in stationary phase (12 h) (**Figure 3.4A**). From these samples, RNA and protein were extracted by standard methods and further characterized by RNA-seq and proteomics analysis, respectively, in duplicates. Differentially expressed transcripts were identified using DESeq2 with a model that included strain, time, and the interaction of these two. From this analysis, 349 transcripts were significantly affected in a Zms4-dependent manner over time ($p_{adj} < 0.01$) and 180 transcripts were affected in a Zms6-dependent manner over time ($p_{adj} < 0.01$) (**Table A.4**). Likewise, protein levels in the Zms4 and Zms6 overexpression strains were also quantified in triplicates by mass spectrometry and normalized by the total spectral counts of each sample to allow for comparison. Those with at least 2-fold increase or decrease upon induction of Zms4 or Zms6 (5 h vs 4 h time points) are listed in **Table A.5**.

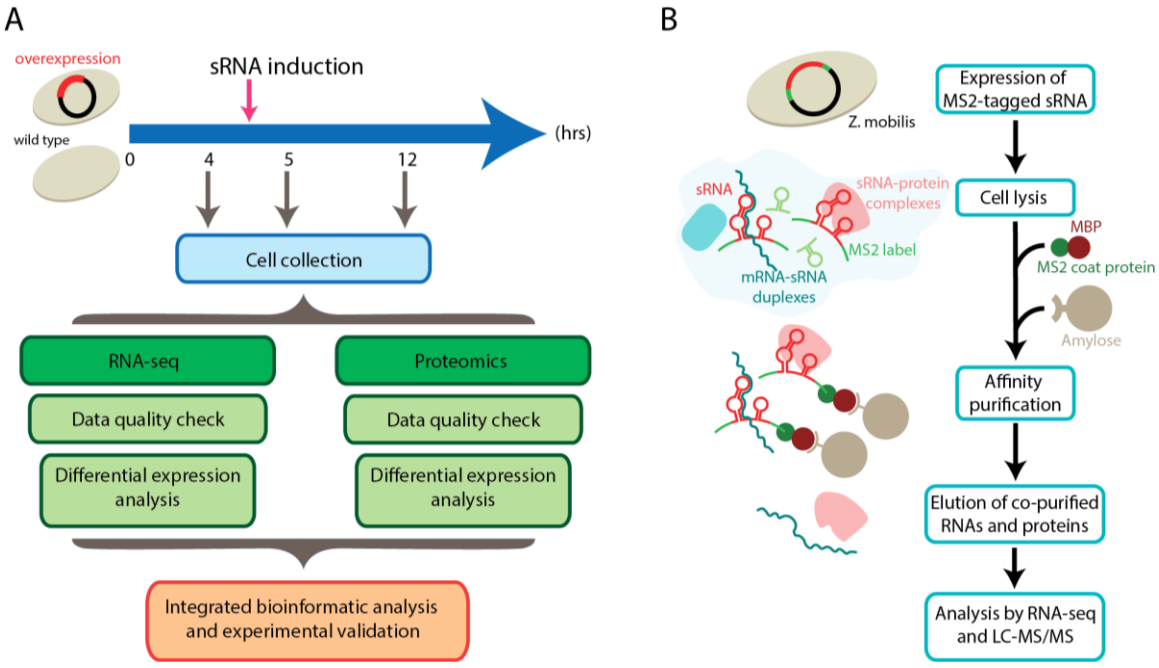


Figure 3.4: Experimental approaches designed to discern gene networks associated with Zms4 and Zms6.

(A) Each overexpression strain along with the wild type was grown in 500 mL cultures anaerobically and sRNA plasmid expression was induced at 4.5 h. Portions of cells were collected at 4, 5, and 12 h of growth, representing before induction, after induction, and stationary phase, respectively. RNA was sequenced and proteins identified by mass spectrometry to discern impacts of Zms4 and Zms6 levels on cellular networks. Differential expression analysis was performed by DESeq2 for RNA-seq data and by Scaffold4 for proteomics data. By integrating the data, patterns of expression at the RNA and protein levels were used to determine most likely targets of sRNA regulation and to infer their regulatory mechanisms. (B) In vivo sRNA interactions were determined by tagging each sRNA with MS2 binding domains and isolating them by affinity purification with maltose binding protein (MBP) fused to the MS2 coat protein and amylose beads. Upon elution, physically associated transcripts and proteins were identified by RNA-seq and mass spectrometry.

Induction of Zms4 especially influenced transcripts involved in transport and biosynthetic process and induction of Zms6 impacted oxidative stress response, cell motility/flagellum organization, and carbohydrate utilization. Both Zms4 and Zms6 additionally influence transcripts coding for proteins associated with translation, hydrogen

sulfide biosynthetic process, sulfate assimilation, and cysteine biosynthetic process. It should be noted that these pathways represent many of the same basal metabolic function that *collectively*, are negatively impacted by ethanol stress and have therefore been implicated in ethanol toxicity. For instance, proteins significantly down-regulated upon Zms4 induction are linked to mismatch repair, chromosome segregation, and DNA topological change while proteins up-regulated upon Zms6 induction include those related to sulfate assimilation and dTTP and hydrogen sulfide biosynthetic processes (**Table A.5**). Collectively, these results show that a considerable portion of cellular transcripts and proteins (directly or indirectly) affected by Zms4 and Zms6 expression levels are involved in ethanol tolerance, particularly through mechanisms that stimulate oxidative stress response, anabolism and DNA repair. Several transcripts and proteins potentially regulated by Zms4 and Zms6 have been previously reported to be differentially expressed under ethanol stress (**Table 3.1**), including some that belong to metabolic and oxidoreductase processes. Given that Zms4 and Zms6 are naturally up-regulated upon ethanol stress, these observations evidence that many transcripts and proteins associated with cellular levels of Zms4 and Zms6 are tightly linked to ethanol tolerance pathways (as well as to other disparate pathways that might not be unique to the stress response).

Table 3.1: List of transcripts and proteins differentially regulated by both ethanol stress and sRNA overexpression in *Z. mobilis*.

ID	Description	GO terms
Transcripts upregulated under both 5% ethanol stress [63] and Zms4 overexpression		
ZMO1025	Anaerobic ribonucleoside triphosphate reductase	ATP binding, DNA replication
ZMO1522	TonB-dependent receptor	Receptor activity, transport
ZMO1647	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	Kinase activity
ZMO2018	Major facilitator superfamily protein	Integral component of membrane
ZMO2030	Hypothetical protein	
Transcripts upregulated under both 5% ethanol stress [63] and Zms6 overexpression		
ZMO1025	Anaerobic ribonucleoside triphosphate reductase	ATP binding, DNA replication
ZMO1522	TonB-dependent receptor	Receptor activity, transport
ZMO2030	Hypothetical protein	
Proteins differentially expressed under both 6% (v/v) ethanol stress [52] and Zms4 overexpression		
ZMO0442	HAD-superfamily hydrolase	Hydrolase activity, metabolic process
ZMO0760	Lactoylglutathione lyase	Dioxygenase activity
ZMO0994	Conserved hypothetical protein	
ZMO1294	Sugar isomerase	Carbohydrate metabolic process
ZMO1334	YceI family protein	
ZMO1544	Cobaltochelatase	ATP binding, cobaltochelatase activity
Proteins differentially expressed under both 6% (v/v) ethanol stress [52] and Zms6 overexpression		
ZMO0442	HAD-superfamily hydrolase	Hydrolase activity
ZMO1712	FKBP-type peptidyl-prolyl cis-trans isomerase	Peptidyl-prolyl cis-trans isomerase activity

3.3.3 Identification of Potential Direct Targets of Zms4 and Zms6

To start identifying specific transcripts and proteins that *directly* interact in vivo with Zms4 and Zms6, we conducted a genome-wide MAPS approach (**Figure 3.4B**) [116]. For these experiments, each sRNA was N-terminally tagged with two MS2-binding domains (sequence in **Table A.6**) and constitutively expressed in the wild type strain under the P_{gap} promoter. Bound complexes were collected during late exponential phase (after 12 hours of growth) and isolated using maltose binding protein (MBP) columns. RNA and protein were extracted by standard methods. Transcripts that pulled down in complex with each sRNA were deep-sequenced and proteins that pulled down with each sRNA were identified by LC-MS/MS. Transcripts at least 1.5-fold more represented in the pMS2-Zms4 and pMS2-Zms6 strains compared to the pMS2 control (lacking sRNA expression) were analyzed further as potential sRNA targets (**Table A.7**). The heat maps in **Figure 3.5** show the behavior of transcripts pulled down by MS2-MBP-sRNA upon induction of Zms4 and Zms6, respectively. Based on the observed expression patterns, potential mechanisms of regulation were hypothesized (**Figure 3.5**). Zms4 was hypothesized to negatively regulate 33 transcripts by inducing transcript degradation or blocking the ribosome binding site (RBS) sequence and to positively regulate 88 genes by transcript stabilization or exposure of the RBS. Similarly, Zms6 was hypothesized to up-regulate 16 and down-regulate 20 transcripts by the same corresponding mechanisms.

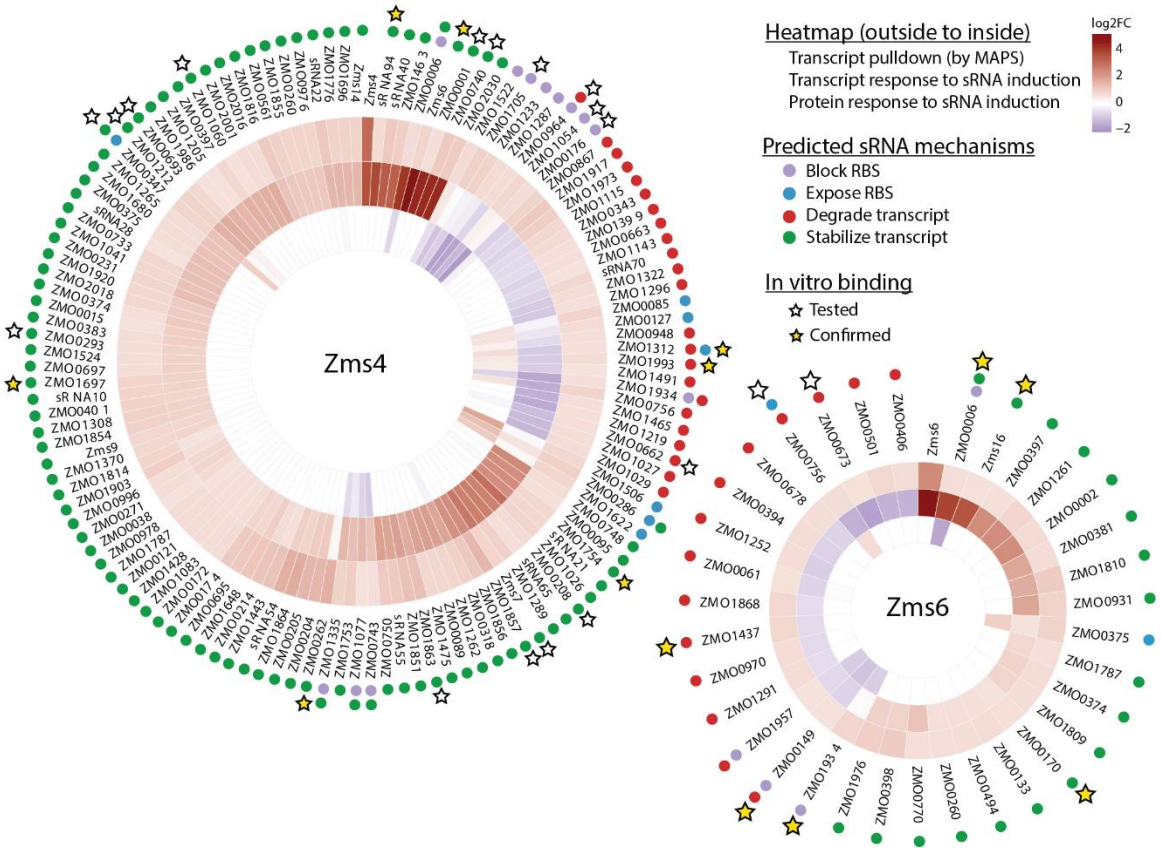


Figure 3.5: Integrated omics analyses reveal most likely targets of Zms4 and Zms6 regulation.

Each heat map shows the genes that co-immunoprecipitated with Zms4 and Zms6 and also showed patterns of transcript and protein expression consistent with canonical sRNA mechanisms: blocking or exposing the RBS and degrading or stabilizing the transcript (colored dots). Heat map tiles (outside to inside) show the log2 fold changes of the sRNA-immunoprecipitation enrichment compared to the control, the transcript expression level at 5 h (after sRNA induction) vs 4 h (before induction), and the protein expression level after vs before sRNA induction. Multiple potential targets with a variety of predicted mechanisms were tested for direct binding with each sRNA in vitro (stars) and several were confirmed (yellow stars).

Although there are only a few well-studied cases of sRNA-protein interactions, we identified 20 proteins (13 non-ribosomal proteins) significantly enriched by co-precipitation with Zms4 and one protein (uncharacterized protein ZMO1657) with Zms6

($p < 0.05$ Fisher's Exact test, Hochberg-Benjamini correction, **Table A.8**). We attribute the pulldown of ribosomal proteins to unspecific binding given their abundance, as previously cited [117]. Of the potential protein targets of Zms4, ZMO0994 (uncharacterized), ZMO1305 (uncharacterized), and ZMO0856 (exodeoxyribonuclease 7 small subunit) also showed at least 2-fold change in expression upon Zms4 induction. Of the proteins co-precipitating with these sRNAs, only phosphoglycerate kinase (ZMO0178) has an annotated nucleotide-binding domain, which is its ATP binding site. There is no common protein pulled out with both Zms4 and Zms6, suggesting that there are no likely RNA chaperones bound. Hfq has been shown important for stress tolerance phenotypes in *Z. mobilis* [83], but was not detected here to co-precipitate with these sRNAs, leaving its functional role unknown in this organism. It may be that sRNAs and their mRNA targets can efficiently interact without an RNA chaperone like Hfq or that some proteins attached in vivo were lost in our purification procedure or difficult to detect by mass spectrometry.

Importantly, one of our most interesting observations from this data is that the Zms4 and Zms6 pulldowns were enriched for several other sRNAs (previously identified to naturally respond to ethanol stress in *Z. mobilis* [53]; this indicated the possibility that Zms4 and Zms6 seed a multi-sRNA network. To our knowledge, this represented the first potential regulatory network constituted by several sRNAs co-interacting in bacteria.

3.3.4 Validation of Direct Targets by EMSA

Given the non-specific interactions that can be captured via pull-down approaches like MAPS (which are likely amplified by the sensitivity of RNA-seq approaches), we next validated physical interactions of the predicted mRNA targets and the corresponding sRNA. We selected target candidates for in vitro testing (marked with a star in **Figure 3.5**) based on their potential relevance to ethanol tolerance (as judged by previous omics studies

in *Z. mobilis* [52,63]) and their likelihood of being fully in *trans* (e.g. their 5' UTRs were not significantly overlapping with neighboring gene coding regions). For these assays, the 5' region (-200 to +100 nt) of each mRNA target was tested for binding with the full sRNA transcript (as previously characterized by RACE analysis [53]) using electrophoretic mobility shift assays (EMSAs). We selected these regions given that most sRNAs are found to base pair within the 5' UTR region of the target they regulate [12]. We confirmed interactions for 7 of the 23 RNAs tested as potential Zms4 targets and for 6 of the 11 RNAs tested as potential Zms6 targets (confirmed interactions in **Figure 3.6A** and **3.6B** and **Table 3.2**; all tested pairs in **Figure 3.7**).

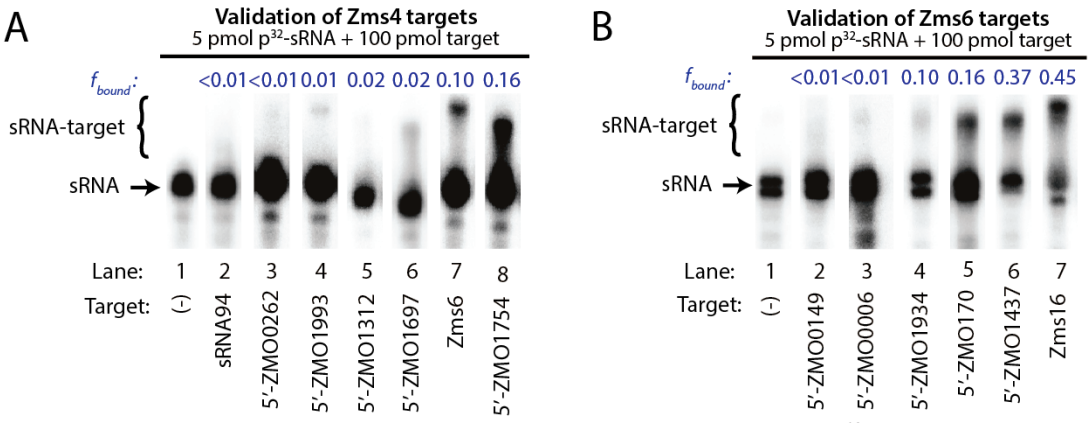


Figure 3.6: sRNA-target interacting pairs verified by EMSA.

Transcripts confirmed to bind in vitro with Zms4 (A) and Zms6 (B) have a range of apparent binding affinities represented by the fraction bound (fbound), calculated from the ratio of the radioisotope signal of the duplex and the total signal of the labeled transcript.

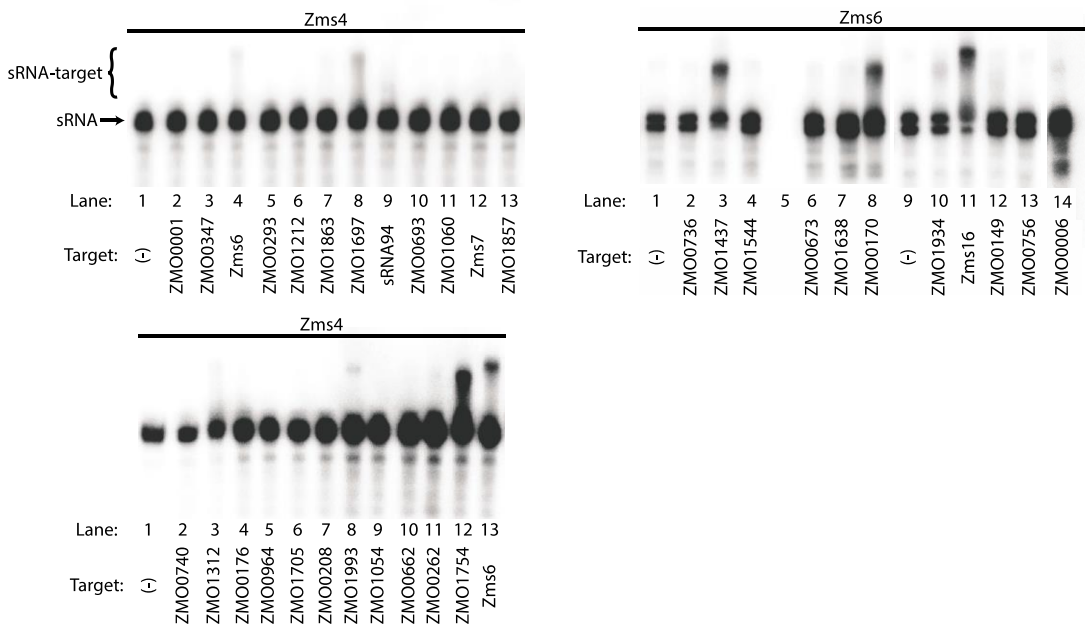


Figure 3.7: Detection of Zms4 and Zms6 in vitro interactions with RNA targets by EMSA.

All tested pairs are shown here and those with a band shift (indicating a confirmed in vitro interaction) are also shown in **Figure 3.6**.

Table 3.2: Confirmed targets of Zms4 and Zms6 regulation.

sRNA	ID	Description	GO terms
Zms4	ZMO1697	Transcriptional regulator, HxIR family	Regulation of transcription
	ZMO1754	Aldehyde dehydrogenase	Oxidoreductase activity
	ZMO1993	Alcohol dehydrogenase GroES domain protein	Oxidoreductase activity; zinc ion binding
	ZMO0262	Heme exporter protein B	integral component of membrane; plasma membrane; heme transporter activity; cytochrome complex assembly
	ZMO1312	Peptidylprolyl isomerase	peptidyl-prolyl cis-trans isomerase activity
	Zms6	sRNA	
	sRNA94	predicted sRNA	
Zms6	ZMO0006	Siroheme synthase	NAD binding; precorrin-2 dehydrogenase activity; sirohydrochlorin ferrochelatase activity; uroporphyrin-III C-methyltransferase activity; cobalamin biosynthetic process; siroheme biosynthetic process
	ZMO0149	tRNA (guanine-N(7)-)-methyltransferase	tRNA (guanine-N7-)-methyltransferase activity
	ZMO1934	N-6 DNA methylase	DNA binding; N-methyltransferase activity; DNA methylation
	ZMO0170	Glycerophosphoryl diester phosphodiesterase-like protein	glycerophosphodiester phosphodiesterase activity; lipid metabolic process
	ZMO1437	Lysine exporter protein	integral component of membrane; plasma membrane; amino acid transport
	Zms16	sRNA	

As shown in **Figure 3.5**, Zms4 was confirmed to bind in vitro to the 5' regions of transcripts it was predicted to stabilize according to the omics data (sRNA94, Zms6, ZMO1754, ZMO0262, and ZMO1697) as well as two transcripts it was predicted to

degrade (ZMO1312 and ZMO1993) or to translationally block at the RBS (ZMO1312). Likewise, Zms6 also shows binding to targets with multiple predicted mechanisms including transcript stabilization (ZMO0006, Zms16, and ZMO0170), RBS blocking (ZMO0006, ZMO1934, and ZMO0149), and transcripts degradation (ZMO0149 and ZMO1437). These results suggest that Zms4 and Zms6 have both activation and repression regulatory capabilities, an interesting observation given the more rare activation role that has been assigned to sRNAs from studies in model bacteria. As seen in **Figure 3.6A** and **3.6B**, the sRNAs interact with variable degrees of affinity with their RNA targets. We found that Zms6 possess higher target affinity than Zms4 by computing the fraction of duplex formed, measured as the ratio of the radioisotope signal of the duplex and the total signal of the labeled transcript.

To identify potential binding sites for the sRNA-RNA interacting pairs, IntaRNA [110] was used. The most favorably predicted binding site locations were mapped to each sRNA's secondary structure inferred from NUPACK [118]. As shown in **Figure 3.8A** and **3.8B**, Zms4 and Zms6 contain multiple functional sites that potentially contribute to multi-tasking the regulation of its targets. Moreover, these sites are predicted to occupy regions of high and low hybridization efficacy, as identified by the InTherAcc biophysical model [119], which could explain the observed differences in binding affinities.

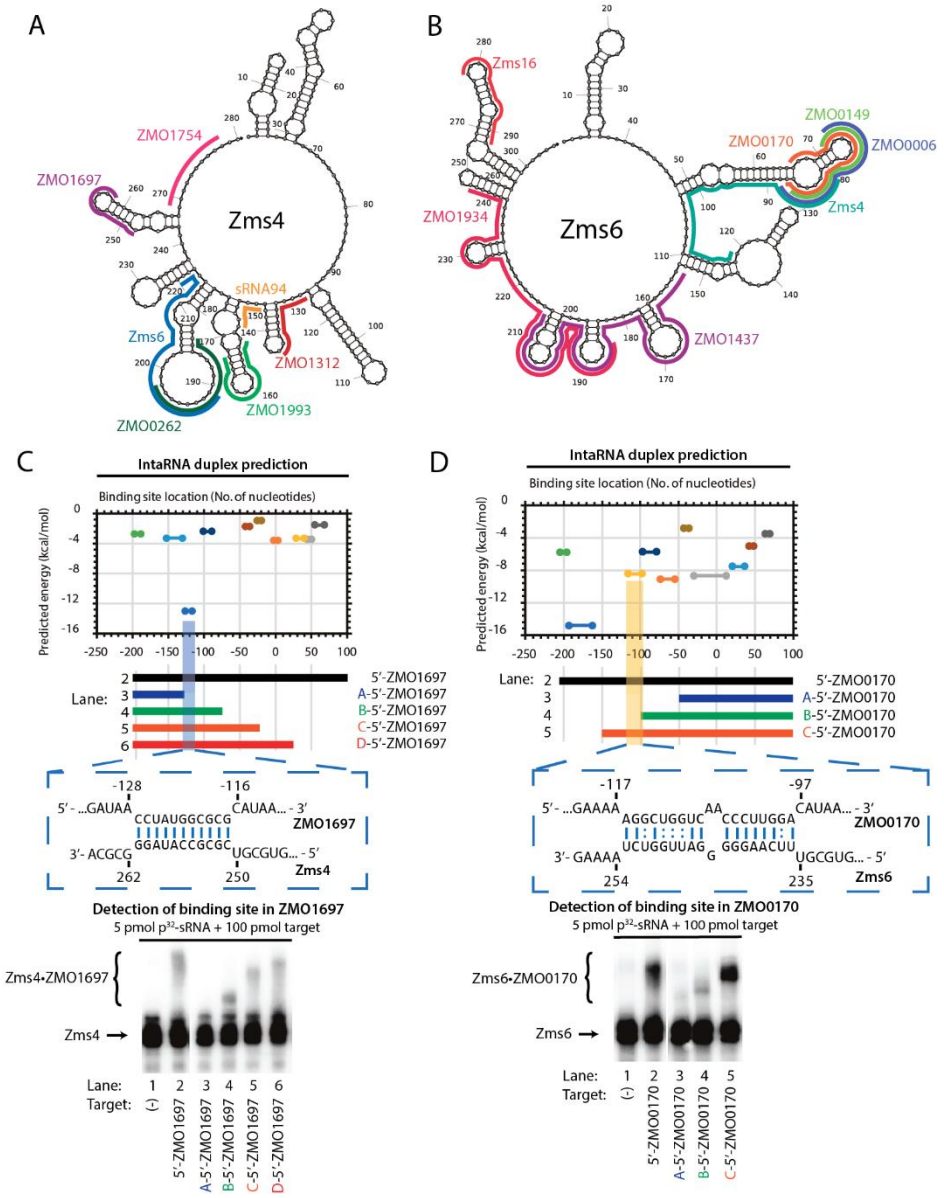


Figure 3.8: Identification of Zms4 and Zms6 binding sites with their mRNA targets.

Binding site locations of these confirmed targets of Zms4 and Zms6 were predicted by IntaRNA and shown mapped onto the secondary structure of sRNAs inferred by NUPACK (A & B). The predicted sites shown represent those with the lowest IntaRNA energy score except for the ZMO0170 site, which was experimentally identified. The predicted binding sequences of the duplex Zms4 • ZMO1697 (C) and Zms6 • ZMO0170 (D) were investigated by EMSA. Different transcript fragments were analyzed and compared with fragments excluding the predicted site to determine the actual sites (highlighted in blue for Zms4 (C) and yellow for Zms6 (D)).

We then experimentally mapped the actual binding site locations in the of two RNA targets by EMSAs. We selected one confirmed ethanol-relevant mRNA target for each sRNA that displayed strong affinity by EMSA band shifting. For Zms4, we studied ZMO1697, which encodes an ortholog of a transcriptional regulator that positively regulates the hxlAB operon associated with detoxification of one-carbon compounds such as methane, methanol, and formaldehyde in *Bacillus subtilis* [120]. For Zms6, we studied ZMO0170 a glycerophosphodiester phosphodiesterase-like protein that hydrolyzes deacylated phospholipids to alcohols and phosphoric esters of glycerol. As exemplified in **Figure 3.8C**, the shortest fragment expanding from -200 to -135 nt in the 5' region of ZMO1697 (labeled as A-5'-ZMO1697), does not exhibit a band shift. The subsequent fragment, B-5'-ZMO1697 (-200 to -78 nt), displays a band shift in agreement with the most optimal binding site predicted by IntaRNA. The next two larger fragments, C-5'-ZMO1697 (-200 to -27 nt) and D-5'-ZMO1697 (-200 to +21 nt), consistently exhibit band shifting with comparable signal indicating that the interactions possibly involve a single binding site 112 nt upstream of the start codon. Through similar testing of Zms6 interacting target ZMO0170 fragments, its binding site was also identified. Specifically, these results show a faint shift in the shorter fragment A-5'-ZMO0170 (-50 to +100 nt). The next larger fragment, B-5'-ZMO0170 (-100 to +100), exhibits a band shift with stronger signal but significant weaker intensity than the signal intensity resulting from the largest fragment C-5'-ZMO0170 (-150 to +100 nt). Taken together, these data revealed that Zms6 is capable of interacting with ZMO0170 in multiple regions but mostly through weak affinity interactions. However, the strong signal in the C-5'-ZMO0170 shift suggests the localization of a dominating site in the region from -150 to -100 nt. By inspection of the IntaRNA predictions, one predicted substructure is confined in this area, ranked fourth in the IntaRNA energy score. All other predicted binding sites for direct targets of Zms4 and

Zms6 are included in **Figure 3.9**. Importantly, these results suggest that the binding sites are located in traditional 5'UTR regions or near the start codon in the coding region.

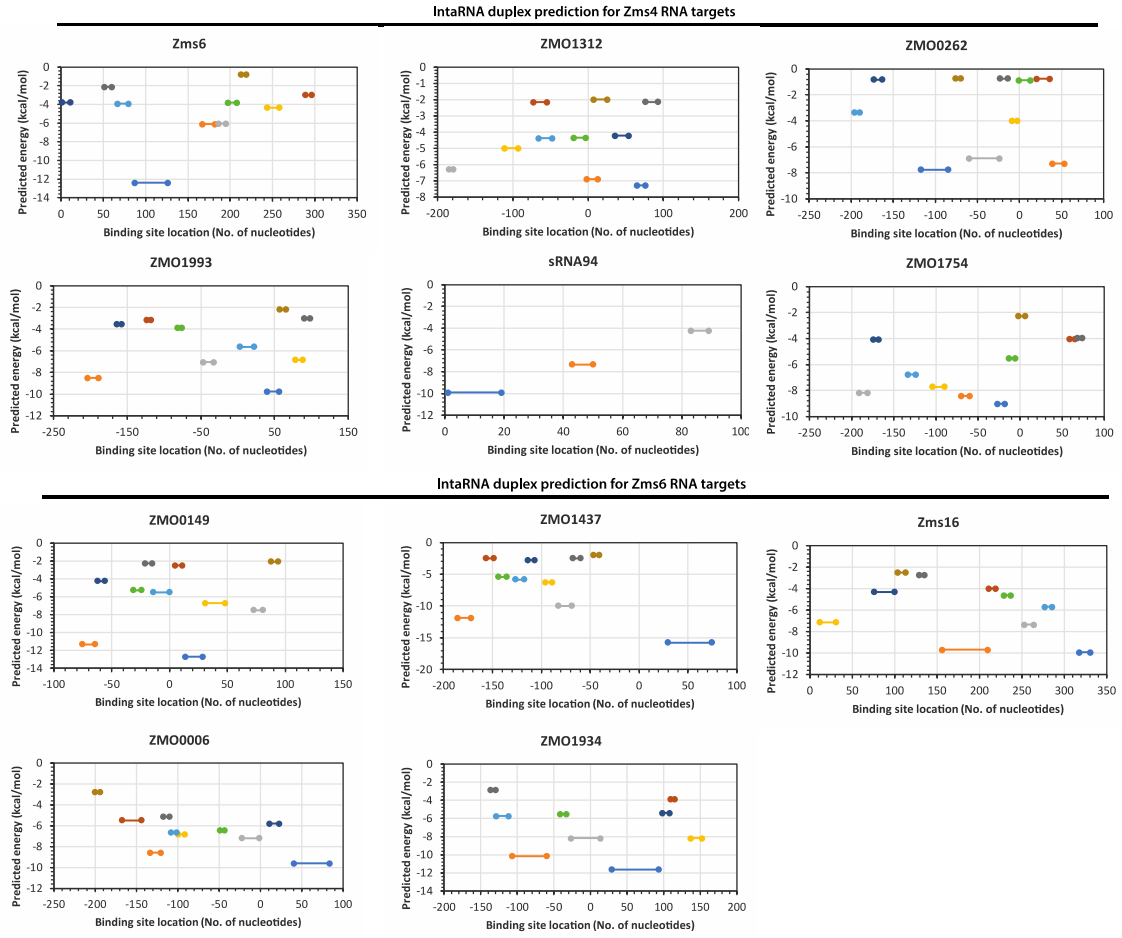


Figure 3.9: Predicted binding sites for direct targets of Zms4 and Zms6 by IntaRNA. Binding sites with the lowest predicted energies are most favorable.

Furthermore, after noting that that various targets could compete for the same binding site on regions of high hybridization efficacy in the IntaRNA predictions, we conducted a competition binding assay using two RNA targets of Zms6, ZMO1437 (an

mRNA encoding for lysine exporter LysE) and Zms16 (another recently uncovered sRNA) since these two RNAs have different predicted binding regions in Zms6. In this experiment, we internally UTP-radiolabeled ZMO1437 and allowed interaction with Zms6 in reactions containing serial increased concentrations of unlabeled Zms16. Interestingly, we observed that the hybridization of a target does not restrict the sRNA to simultaneously interact with another RNA on a second binding site (**Figure 3.10**).

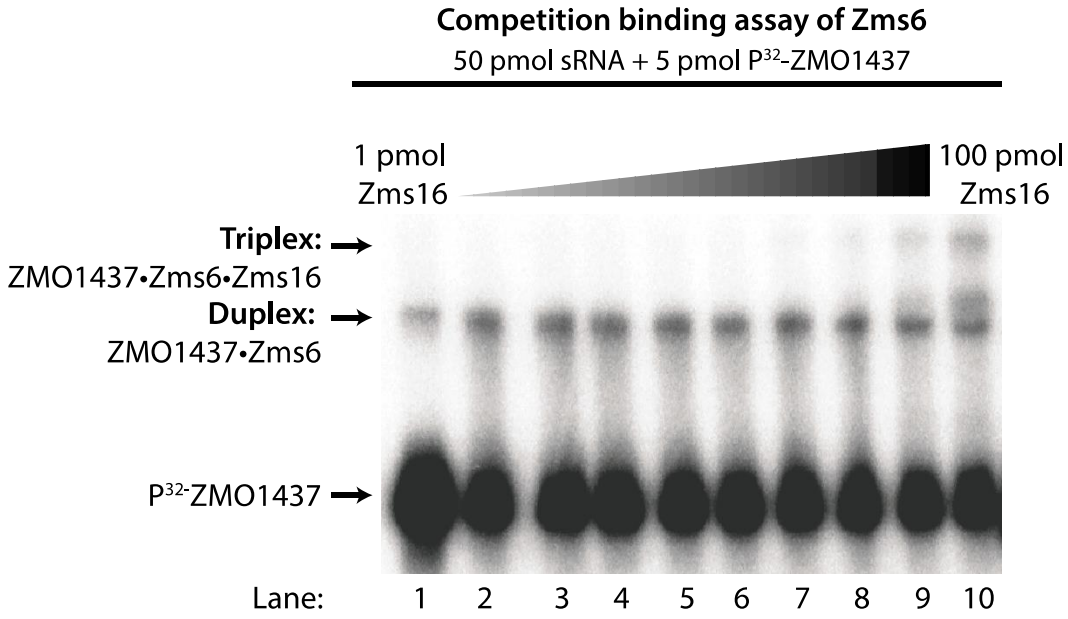


Figure 3.10: Competition binding assay shows simultaneous target binding.

5 pmol of ^{32}P-labeled ZMO1437 was allowed to interact with 50 pmol of unlabeled Zms6, and compete with unlabeled Zms16. Zms16 was serially titrated in the binding reactions by 2-fold increments.

3.3.5 Combinatorial Effects of sRNAs on Ethanol Tolerance Shows Complex Effect

Given the lack of overlap between the confirmed set of mRNA targets of Zms4 and Zms6 and the fact that these participate in important pathways associated with alcohol and chemical tolerance (**Table 3.2**), we hypothesized that Zms4 and Zms6 could mediate crosstalk between these pathways. In this way, these two sRNAs can link regulation of a broader set of pathways, establishing a wider regulatory network of relevance to a complex phenotype like ethanol tolerance. Considering the interactive network of multiple sRNAs that appear seeded by Zms4 and Zsm6 (e.g. observed EMSA interactions of Zms4 to Zms6 and sRNA94; Zms6 to Zms4 and Zms16), we further investigated the combinatorial effect of sRNAs on ethanol tolerance. For these experiments, strains overexpressing each possible combination pairs of Zms4, Zms6, and Zms16 were developed, as well as a strain expressing all three. As **Figure 3.11** shows, most of the combination strains have growth rates significantly higher than the wild type and empty plasmid strains in 8% (v/v) ethanol, although, comparable to the individual sRNA overexpression strains in 8% (v/v) ethanol (compare **Figure 3.1A** and **Figure 3.11**). Yet, the strain combination of overexpression of Zms6 and Zms16 (pZms6-16 strain) shows the highest growth rate of all the strains tested under ethanol stress. These results confirmed the importance of Zms6 to growth on the ethanol stress, suggesting the possibility that growth rate increases observed by overexpressing Zms6 might be close to the limit of what can be achieved. Interestingly, the pZms4-6-16 strain shows the slowest growth rate, likely due to the cellular overload and major metabolic disruption of harboring a plasmid overexpressing three different sRNAs (albeit under the same P_{tet} promoter). Overall, these results show that the benefits of sRNA expression are not directly additive in this system as previously reported for three sRNAs involved in *E. coli* acid tolerance [27], perhaps related to the physical interactions of the

sRNAs observed in this study. This network represents a potentially delicate interplay of sRNAs defined by optimal stoichiometric ratios.

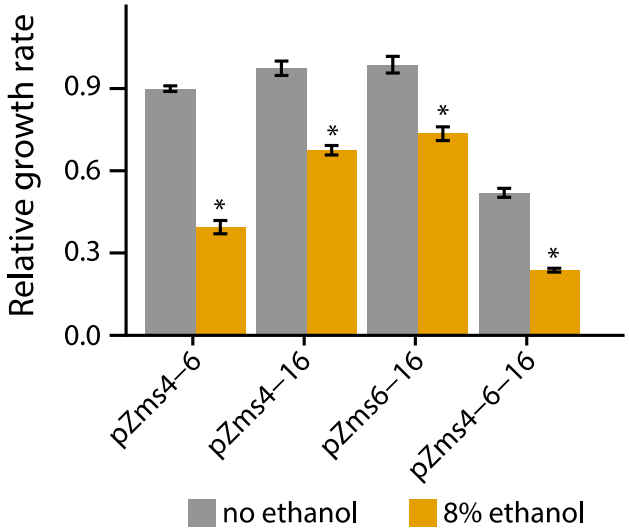


Figure 3.11: Combined sRNA overexpression indicates complexity of the network.

Average growth rates were calculated for each sRNA overexpression and deletion strain with and without 8% (v/v) ethanol supplementation grown in 300 µL cultures in the Bioscreen C. Values represent mean ± SD of biological triplicates, normalized to the wild type growth rate in the no ethanol condition (**Figure 3.1A**). Significant values are *P ≤ 0.001 from Student's t-test of each strain compared to pEmpty (**Figure 3.11A**).

3.4 Discussion

An abundance of sRNAs has been discovered in numerous species, but characterization of these powerful regulators (particularly their interacting target pairs) lags far behind, limiting our understanding of how sRNAs control complex cellular phenotypes. In this study, two sRNAs that are naturally differentially expressed under ethanol stress in *Z. mobilis* are shown to be key to ethanol tolerance and shown to coordinate an entire network of gene regulation that includes other sRNAs. Zms4 and Zms6 represent the first sRNAs with regulatory functions confirmed in *Z. mobilis*. We showed that without these

sRNAs, cells are highly sensitive to ethanol stress, and that by manipulating their cellular levels, ethanol tolerance and production can be improved.

In this study, multi-omics analyses (for instance, integrative approaches like INFO [96]) played a key role in uncovering the network of sRNA's and their interacting target pairs. Especially insightful in this study was the use of patterns in the omics data to infer mechanisms of sRNA action on potential targets (**Figure 3.5**). Using these methods, multiple sRNA-interacting pairs were identified for both Zms4 and Zms6 in this study with experimental confirmation of multiple specific binding sites throughout each sRNA (**Figure 3.8A** and **3.8B**). Interestingly, all but one of the predicted sites corresponds to stem-loops in the sRNA secondary structure. RNA-RNA binding site interactions can be difficult to predict computationally due to the sensitivity of these interactions to structural complexity and in vivo variables beyond current modeling capabilities. In this study, IntaRNA appeared to be effective in identifying true binding site locations. The Zms4-ZMO1697 binding site confirmed by EMSA matched the location predicted as most favorable by IntaRNA. For the Zms6-ZMO0170 interaction, our EMSAs showed multiple potential binding sites, but the dominating one was predicted in the top five most favorable IntaRNA predictions. Our competition assay in which two Zms6 targets were observed to bind simultaneously also supports the prediction of their separate binding sites by IntaRNA. The use of IntaRNA to predict sRNA-mRNA pairs genome-wide was less successful as of the 10 sRNA-mRNA pairs confirmed in this study, only 2 were in the top 100 IntaRNA predicted Zms4- or Zms6-mRNA binding pairs.

A schematic of the sRNA network elucidated in this study is shown in **Figure 3.12**. Through sRNA co-immunoprecipitation, we identified transcripts and proteins that physically associated with Zms4 and Zms6 and that displayed changes in their expression patterns in Zms4 overexpression strains. Zms4 has three mRNA targets involved in

oxidoreductase activity including HxIR family transcriptional regulator (ZMO1696), aldehyde dehydrogenase (ZMO1754), and alcohol dehydrogenase (ZMO1993). Both Zms4 and Zms6 have one confirmed mRNA target involved in transport, heme exporter (ZMO0262) and lysine exporter (ZMO1437). Heme exporters have been implicated in acid tolerance in *Shewanella oneidensis* [121] and resistance to oxidative stress in *E. coli* [122]. This lysine exporter in *Z. mobilis* (ZMO1437) has been previously identified to contribute to tolerance to phenolic aldehydes, suggesting it could play a role in multiple stress response pathways [123]. Most importantly, these newly characterized sRNAs co-regulate these key pathways. The concept of a network of sRNAs that each regulates multiple mRNA targets as seen here in *Z. mobilis* is novel and there are limited reports of direct sRNA-sRNA interactions in any context. To our knowledge, the interactions of Zms4, Zms6, Zms16, and sRNA94 in *Z. mobilis* therefore form the largest known sRNA-sRNA interaction network. Based on the upregulation upon Zms4 induction, it appears that Zms4 may stabilize Zms6 and sRNA94 transcripts. Similarly, Zms6 appears to stabilize Zms16. These results suggest a synergy between these sRNAs whereas the network in that the expression of one may enhance the regulatory impact of others on the corresponding targets and pathways.

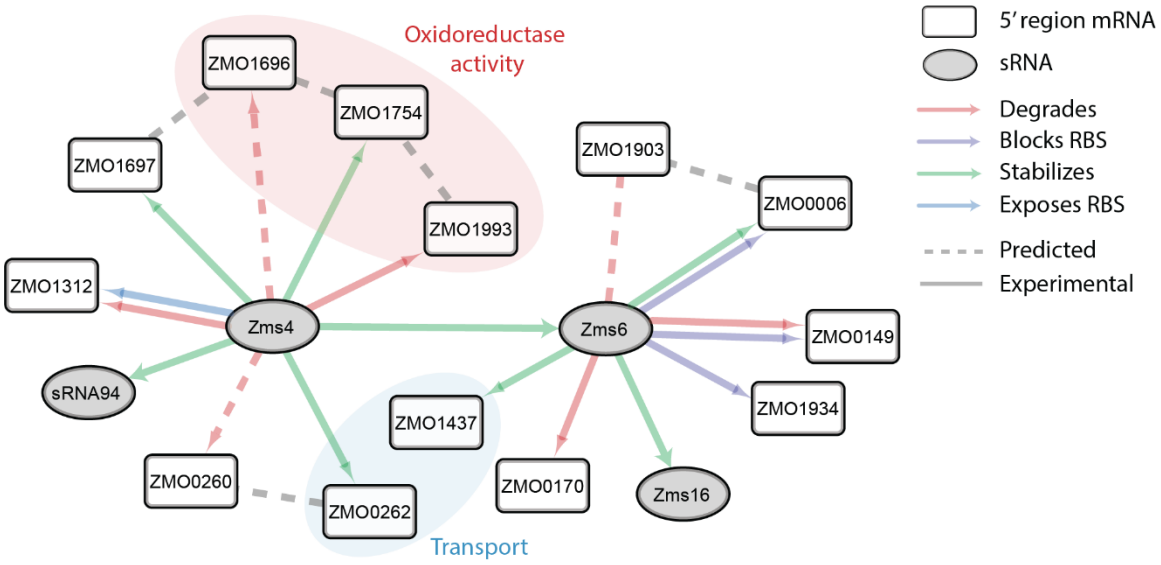


Figure 3.12: A network of sRNA-target interacting pairs were identified in this study. Confirmed in vitro binding pairs are illustrated by solid arrows with colors indicating the predicted sRNA mechanism according to the patterns of integrated omics data (**Figure 3.5**). Dashed lines are protein-protein interactions predicted by STRING [124].

While patterns of expression in the omics analysis post sRNA induction allowed us to narrow the pool of potential direct targets and to hypothesize their mechanisms, it is possible that targets not detected to bind in vitro are either highly very dependent on the in vivo environment and/or bind to other locations of the mRNAs, outside the 5' regions tested here. Interactions between proteins of the mRNA targets as predicted by STRING [124] were also mapped in **Figure 3.12** to illustrate potential cross talk between these. Note that only Zms4 and Zms6 were specifically characterized in this study, so additional pathways may be affected given the likelihood that Zms16 and sRNA94 have their own separate targets. Additionally, because ZMO1697 is a transcriptional regulator, additional downstream impacts on gene expression would take effect through it based on Zms4 regulation.

Although the inferred mechanisms of Zms4 and Zms6 regulation on their targets were not specifically validated in this study, for several targets, these mechanisms are consistent with those observed in other stress responses in *Z. mobilis* and in other organisms. For instance, HxIR transcriptional regulator (ZMO1697) transcript has been shown to be up-regulated in ethanol stress [63], in adaptation to high glucose concentrations [125], and in *B. subtilis* formaldehyde stress. Based on our omics data, it appears that Zms4 stabilizes this transcript, suggesting that Zms4 may be involved in the up-regulation of this transcript's expression in ethanol and other stresses. Likewise, both the transcript and protein of alcohol dehydrogenase (ZMO1993) were down-regulated in the acetate tolerant mutant AcR strain compared to the ZM4 strain [51]. This result is consistent with our predicted mechanism of Zms4 regulation, which is to degrade the ZMO1993 transcript, thereby also decreasing protein expression. In *Clostridium thermocellum*, a mutation to alcohol dehydrogenase improved ethanol tolerance through changing NADH-dependent activity to NADPH-dependent, thus altering the electron transport chain in the mutant [126]. Alcohol dehydrogenase activity has also been positively correlated with ethanol tolerance in *Drosophila* [127]. Aldehyde dehydrogenase (ZMO1754) transcript is expressed 8-fold higher in xylose-only media compared with glucose-only media [64]. It appears the Zms4 may be responsible for stabilizing this transcript in ethanol stress. In humans, both coding and noncoding variations of alcohol and aldehyde dehydrogenase enzymes influence alcohol metabolism and therefore risk of alcoholism [128]. Lysine exporter (ZMO1437) transcript is up-regulated in stress response of multiple phenolic compounds [123] and Zms6 potentially stabilizes this transcript according to our data. Taken together, these results suggest that Zms4 and Zms6 may be part of general stress response mechanisms that largely contribute to cell survival under ethanol stress. Regulatory RNA in other organisms may serve similar functions in coordinating network responses to alcohol stress.

Approaches to understand and engineer alcohol tolerance should not overlook regulatory RNAs.

Chapter 4: A Bioinformatics Pipeline Developed to Identify Relevant Regulatory Small RNAs for Synthetic Phenotype Construction

4.1 Chapter Summary

As global controllers of gene expression, small RNAs represent powerful tools for engineering complex phenotypes. However, a general challenge prevents the more widespread use of sRNA engineering strategies: mechanistic analysis of these regulators in bacteria lags far behind their high-throughput search and discovery. This makes it difficult to understand how to efficiently identify useful sRNAs to engineer a phenotype of interest. To help address this, we developed a forward systems approach to identify naturally occurring sRNAs relevant to a desired phenotype: RNA-seq Examiner for Phenotype-Informed Network Engineering (REFINE). This pipeline uses existing RNA-seq datasets under different growth conditions. It filters the total transcriptome to locate and rank regulatory-RNA-containing regions that can influence a metabolic phenotype of interest, without previous mechanistic characterization. When applied to *Escherichia coli* iron-starvation stress, the pipeline identified well-characterized iron-stress sRNA regulator RyhB. Additionally, this approach detected sRNAs related to ethanol tolerance in non-model ethanol-producing bacterium *Zymomonas mobilis*. We overexpressed sRNA candidates predicted by REFINE to identify their impact on ethanol tolerance. The overexpression of Zms8, a high scoring sRNA candidate, improved ethanol tolerance. In this way, the REFINE approach informs strain-engineering strategies.

4.2 Introduction

With rising demands of efficiency and sustainability, the use of microbes as chemical factories is increasingly attractive. In bacteria, small RNAs (sRNAs) regulate cellular pathways and metabolic engineers increasingly exploit them for engineering

purposes [17–20,24]. sRNAs regulate mRNA and protein expression, typically by blocking translation or changing stability [12]. Many natural sRNAs respond to environmental signals and coordinate network responses. Their short length (50-300 nt), dynamic nature, and multi-target impacts make them especially attractive for engineering complex phenotypes.

Current engineering efforts to use sRNAs focus primarily on the design of synthetic transcripts to knock down expression of specific mRNA targets, typically by blocking their ribosome binding sites (RBS) [24]. These targeted knockdowns are useful for optimizing individual pathways but are limited in addressing complex phenotypes like stress tolerance, which involve large sets of genes [129].

While strategies for engineering natural sRNAs have been successful, they have been mostly limited to well-characterized pathways in model organisms. For example, the overexpression of naturally-occurring sRNAs RprA, ArcZ, and DsrA improved acid tolerance in *Escherichia coli* by 8500-fold [27]. Similarly, overexpression of sRNA RyhB in *E. coli* increased production of 5-aminolevulinic acid by 16% [130]. Other phenotypes improved by natural sRNA engineering strategies include succinate, fatty acid, amorphadiene, and butanol production [131–133]. In these cases, the wealth of previous sRNA characterization (known mRNA targets and mechanisms) enabled engineers to foresee and achieve phenotype goals [30,134].

A number of existing tools locate sRNAs including QRNA, Intergenic Sequence Inspector, RNAz, sRNApredict/SIPHT, sRNA scanner, and nocoRNAc [135–141]. But, these rely on conservation of sequence and/or structure and depend on the set of known sRNAs and on homology, often lacking in non-model organisms. Additionally, most of these programs are not readily available for current users.

sRNA candidates predicted by these tools require experimental validation as they carry no evidence of actual transcript expression in vivo. From a list of candidates,

researchers typically screen each by northern blot. Considering the low-throughput nature of northern blotting, many sRNA candidates from long lists may go untested. Upon identification of an sRNA with detectable expression, follow-up experiments may include knockout or overexpression to observe phenotype impacts and gel shift assays to check binding with any mRNA targets predicted by programs like IntaRNA [110]. Ultimately, the process is slow and lack direction toward metabolic engineering goals.

Expression-based approaches are more suitable to identify stress-responsive or phenotype-relevant sRNAs [142]. For example, in the initial sRNA discovery efforts in *Z. mobilis* [53], visual inspection of transcriptome data yielded 95 sRNA candidates and detected expression of 15 by northern blot. In this study, sequence-based approaches, WU-BLAST [143] and SIPHT [136] contributed 20 and four sRNA candidates, respectively. Only 10 of the 95 candidates identified by transcriptome data overlapped with the sequence search method sets. Ultimately, the sequence-based tools only contributed two of the 15 sRNAs verified by northern blot. This suggests the sequence-based tools, particularly in non-model organisms, leave many sRNAs undetected. Additionally, sRNAs that contribute to a phenotype unique to an organism (like the high ethanol tolerance of *Z. mobilis*), may not prove widely conserved since the phenotype lacks conservation as well.

sRNA candidates predicted by these tools require experimental validation. Transcriptome approaches do represent in vivo expression, but noise from non-specific read mapping casts some doubt and specific transcript ends prove difficult to discern without follow-up experiments.

Advances in these sequence search algorithms and in high-throughput sequencing enabled the discovery of hundreds of sRNAs across bacteria [106,107], but characterization lags far behind. The vast majority of sRNAs remain without any known function. Mechanistic characterization of these sRNAs requires low-throughput knockout and

overexpression studies, a particular challenge in non-model organisms [144–146]. For metabolic engineers, it has been impractical to consider the large (and growing) pool of sRNA regulators for specific goals since most sRNAs lack foreseeable roles in producing phenotypes of interest.

Meanwhile, RNA-seq datasets accumulate in public databases, containing a wealth of insight into regulatory networks not yet extracted [147]. The growth conditions and phenotypes documented in these studies connect strain performance with RNA and protein expression profiles.

In this paper, we propose a bioinformatics pipeline to identify sRNAs relevant to desired phenotypes using existing "omics" datasets (**Figure 4.1**). This systematic approach does not require previous sRNA characterization or discovery. We aim to exploit transcriptome data to predict regulatory networks and then characterize phenotypic impacts of predicted sRNA regulators. By prioritizing the list of sRNA candidates, we save experimental time and cost.

To demonstrate the value of this new approach, we select industrially relevant phenotypes with known sRNAs involved in their regulation and RNA-seq data available. With our bioinformatics approach, we show a more efficient way to find useful RNA regulators to construct a phenotype of interest, prioritizing the most promising sRNA candidates for any biochemical analysis.

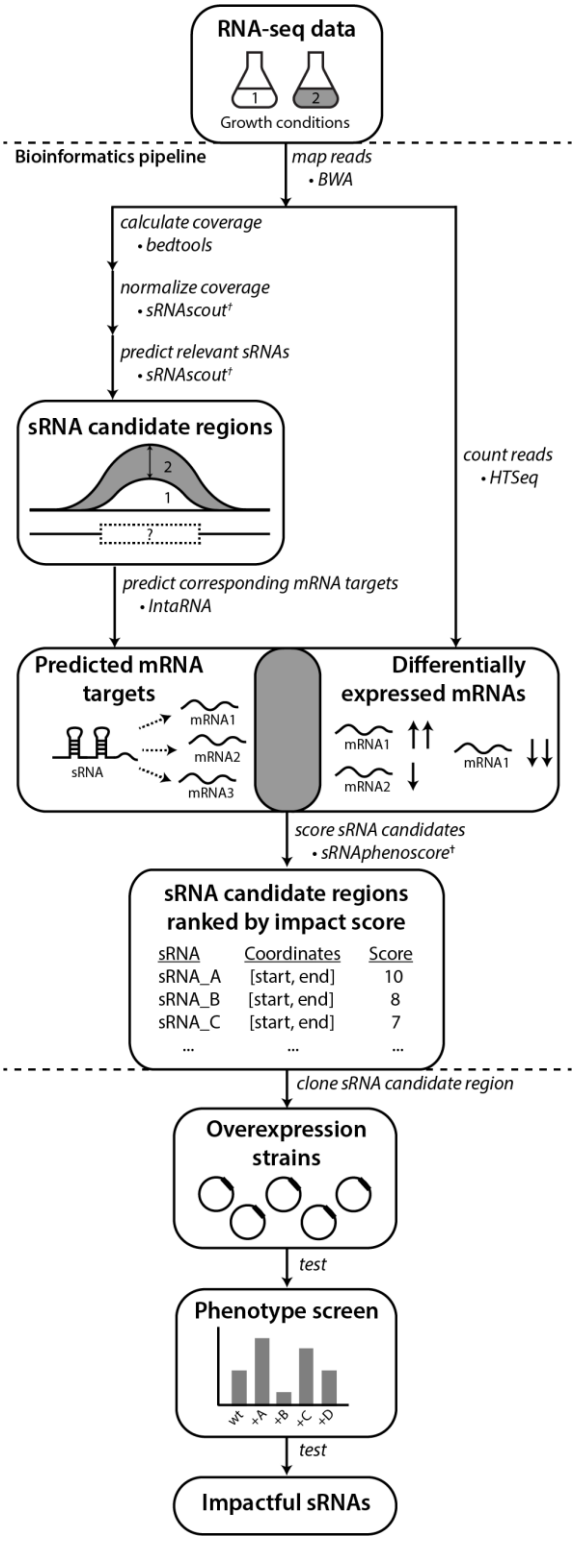


Figure 4.1: REFINE approach identifies sRNAs relevant to phenotypes of industrial interest.

Existing computational tools were combined with new tools (†) developed in this work. Raw RNA-seq data from two conditions related to the phenotype of interest are analyzed for differentially expressed RNAs using standard tools for alignment (BWA) and counting (HTSeq). The REFINE sRNA prediction step identifies differentially expressed intergenic regions likely to contain sRNAs based on their length and a minimum expression level. For each sRNA candidate identified, IntaRNA predicts its most likely mRNA targets based on binding energies. Differential expression of these predicted mRNA targets between the two experimental conditions suggest possible regulation by the sRNA. A score is assigned to each sRNA based on the differential expression of its predicted targets and also the differential expression of the sRNA transcript itself. In this scoring equation, predicted mRNA targets with higher IntaRNA ranking contribute more heavily to the score. The calculation of a total score for each sRNA produces a ranked list of genomic regions most likely to contain sRNAs relevant to the phenotype. These sRNA candidates can each be overexpressed on plasmids and tested for their impact on the phenotype.

4.3 RESULTS

4.3.1 Development of REFINE Approach to Identify sRNAs that Affect Phenotypes of Interest

We developed a bioinformatics pipeline to identify sRNA candidates most likely to impact a phenotype of interest. In this case study, we focus on iron starvation (for *E. coli*) and on ethanol tolerance (for *Z. mobilis*). As shown in **Figure 4.1**, the input to the bioinformatics process is RNA-seq data from two growth conditions, based on the phenotype goal. For example, if the engineering goal were to develop a strain tolerant to a high-temperature industrial process, RNA would be collected and sequenced at a high temperature (stressed condition) and at its normal temperature (unstressed condition). The comparison of these conditions allows observation of how the transcriptome changes naturally as part of the cellular response to the changing condition (or stress). As an underlying design principle of this analysis, we expect that with maximal stress the transcriptome response exhibited is more robust (as long as the imposed stress condition doesn't hinder survival) and therefore the variability in RNA levels becomes significant enough to draw conclusions.

Public databases such as the Sequence Read Archive (SRA) of NCBI contain an abundance of RNA-seq data [148]. RNA-seq data used with the REFINE pipeline should not be depleted of small transcripts during RNA extraction and library preparation (although rRNA depletion is appropriate). To determine which direction sRNAs are encoded in the genome, requires a strand-specific RNA-seq library prep [149]. Considering the noise that can occur in RNA-seq datasets, biological replicates increase the reliability of the output. After experimental data collection, BWA maps the transcriptome data to a reference genome with a standard workflow using, yielding a landscape of transcript read abundance across the genome (**Figure 4.2A**). Then, the calculation of per-nucleotide coverage

represents the number of reads mapped to each genomic location. We normalize coverage by each sample's total read count to allow averaging of replicates and comparison between conditions.

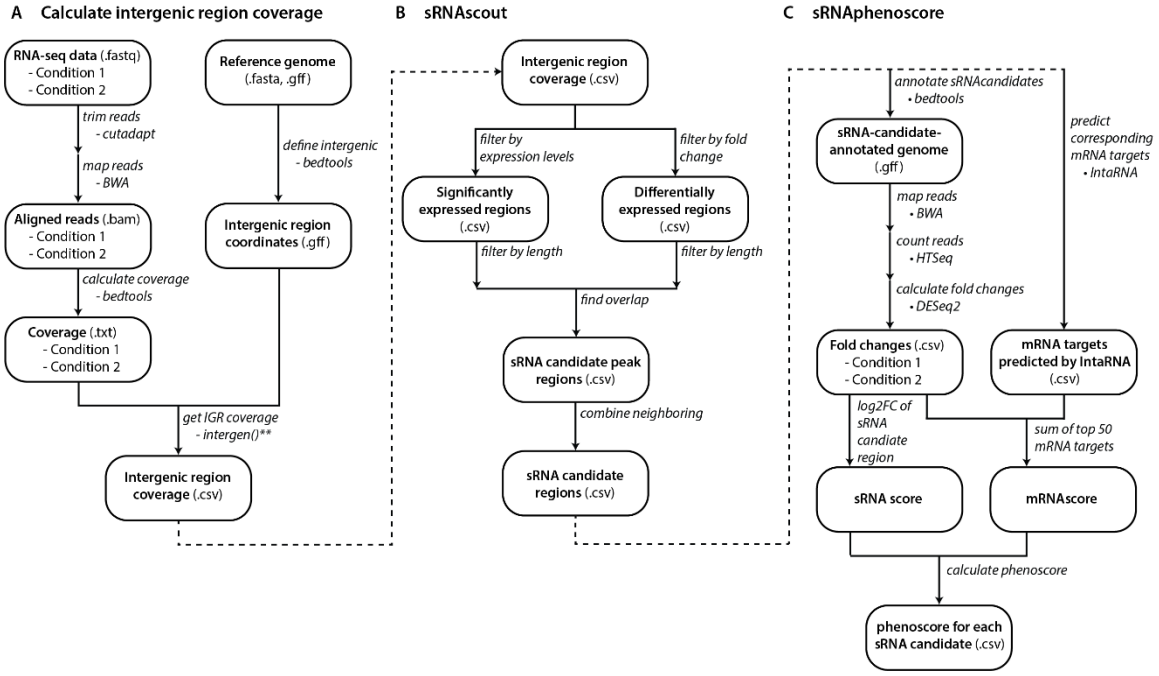


Figure 4.2: Computational workflow from raw RNA-seq data to phenoscores of each predicted sRNA candidate.

(A) With raw RNA-seq data and a reference genome, files with coverage of intergenic regions are made for each growth condition. **(B)** sRNAscout filters intergenic regions down to short peak regions of minimum expression and differential expression between the two growth conditions. Neighboring peaks are combined into single regions that likely represent single transcripts. **(C)** The phenotypic impact of each sRNA candidate is predicted by sRNAphenoscore. The phenoscore is composed of two components: the sRNAscore and the mRNAscore. The sRNAscore is the absolute value of the log2 fold change of expression in the sRNA candidate region between the two growth conditions represented in the RNA-seq data. The mRNAscore is calculated by predicting the top 50 mRNA targets of each sRNA candidate by IntaRNA. It is calculated as the fold changes of each mRNA in the RNA-seq data between the two conditions scaled by its predicted energy from IntaRNA. The sRNAscore and mRNAscores are normalized by the max of each across all sRNA candidates and summed to produce the final phenoscore for each sRNA candidate.

The sRNAscout program (**Figure 4.2B**), a new algorithm developed in this work, uses the normalized per-nucleotide coverage files from each condition to predict genomic regions most likely to contain sRNAs that affect a particular phenotype of interest – we call these sRNA candidate regions. The sRNAscout program replaces the visual inspection and manual scrolling through transcriptome as used in the past. The primary advantage of this approach over other sequence-homology sRNA prediction tools is the targeted nature of the search for sRNAs that respond in vivo to changing environmental conditions (as represented by the two sets of input transcriptome data). Additionally, sRNAscout can identify more unique sRNAs that lack homology to those found in model organisms.

sRNAscout first filters the total transcriptome down to only transcripts mapping to intergenic regions, where RNA- and protein-encoding genes have not been annotated and where regulatory sRNAs have been found abundant. This step depends on a reference genome file including known mRNA, rRNA, tRNA, and any other transcripts that should not be classified as "intergenic" for the sake of sRNA prediction. The reference genome used here must contain the direction of each known transcript along with its genomic coordinates.

Next, sRNAscout searches intergenic regions for sRNA candidate peaks, defined by a sufficient, minimum expression level and differential expression. Previous sRNA identification work identified these criteria as important hallmarks of true sRNAs in [53,150]. sRNAscout combines peak regions within 50 nt of each other into a single sRNA candidate region, since these likely constitute a single transcript. The program expands any short peak regions to 200 nt with the peak as the center.

From these potentially sRNA-containing regions, we aim to discern which most likely regulate genes associated with the target phenotype. We hypothesize an sRNA linked to the phenotype will exhibit two main features. First, we expect the sRNA

expression level to change between the two representative growth conditions, in the presence and absence of the stress of interest. The sRNAscore of the algorithm quantifies this feature as the magnitude of the log2 fold change in expression level between these conditions. Second, we expect the sRNA's potential mRNA targets to show differential expression between the two conditions (in the presence and absence of the stress). The mRNAscore quantifies this.

Specifically, to determine mRNA targets that likely regulated by the sRNA candidates, we employ an RNA-RNA interaction prediction program (IntaRNA) to yield a list of mRNAs in the organism with the most favorable binding energies to the sRNA [151]. Among the available tools for sRNA target prediction, IntaRNA shows the highest accuracy without requiring the input of homologous sRNAs in other organisms [152]. Benchmarking of IntaRNA showed the median rank of experimentally verified targets was 10 [152]. Here, we use the top 50 mRNA targets predicted by IntaRNA for each sRNA candidate to quantify the potential impact of the sRNA on the expression of these mRNA predicted targets. We scale the magnitude of the log2 fold expression change of each predicted mRNA target between the two growth conditions (with and without the stress) by its predicted free energy of binding with the sRNA candidate. The sum of these scaled values for the top 50 mRNA targets represents the mRNAscore.

Finally, the program combines the sRNAscore and mRNAscore into a total phenoscore for each sRNA candidate (**Figure 4.2C**), in which a higher score indicates a region more likely to contain an sRNA that affects the phenotype of interest. As a result, the bioinformatics pipeline produces a list of sRNA candidate regions ranked by their phenoscore.

To experimentally verify the most promising candidates, we generated a library of overexpression strains, each containing an sRNA candidate region, and screened these strains for phenotype impact. We concluded that sRNA overexpression strains with significantly different performance compared to the wild type (or to an empty plasmid control) reveal useful sRNA candidates that can contribute towards engineering of a phenotype of interest. In this way, we expect to inform strain engineering strategies in which we overexpress sRNAs that enhance a particular phenotype and knock down sRNAs that negatively impact the phenotype.

4.3.2 REFINE Approach Identifies Well-characterized Iron Stress Regulator sRNA RyhB in *E. coli*

To test the ability of the REFINE approach to identify impactful sRNAs, we applied it to an low-iron survival phenotype regulated by the well-characterized sRNA, RyhB. The iron-starvation stress response is critical for pathogen survival as they invade their host. RyhB homologs exist in a number of bacterial pathogens [153]. As a 90-nt sRNA with multiple known mRNA targets, RyhB exhibits multiple mechanisms of regulation that control iron-starvation stress in *E. coli*. Out of the 56 suggested targets (by co-immunoprecipitation or RNA-seq of RyhB mutants), biochemical assays defined 12 mRNA targets direct interactions (binding sites confirmed) so far [153]. Importantly, overexpression of RyhB results in increased concentration of free intracellular iron, decreased swarming motility, and increased production of succinate, acetate, and 5-aminolevulinic acid [130,131,154,155]. Deletion of RyhB severely limits siderophore production and eliminates intracellular iron homeostasis [156]. Conveniently, RNA-seq has been performed on *E. coli* K-12 strains with and without iron (minimal media supplemented with $FeCl_2$ or iron chelator 2,2'-dipyridyl, **Figure 4.3A**) to study network responses [157].

Now, the REFINE approach (**Figure 4.1**) can uncover sRNAs such as RyhB from this publicly available data containing biological duplicates (NCBI GEO GSE54901). For these reasons, identification of RyhB in iron regulation phenotype represented an ideal proof of principle study for this work.

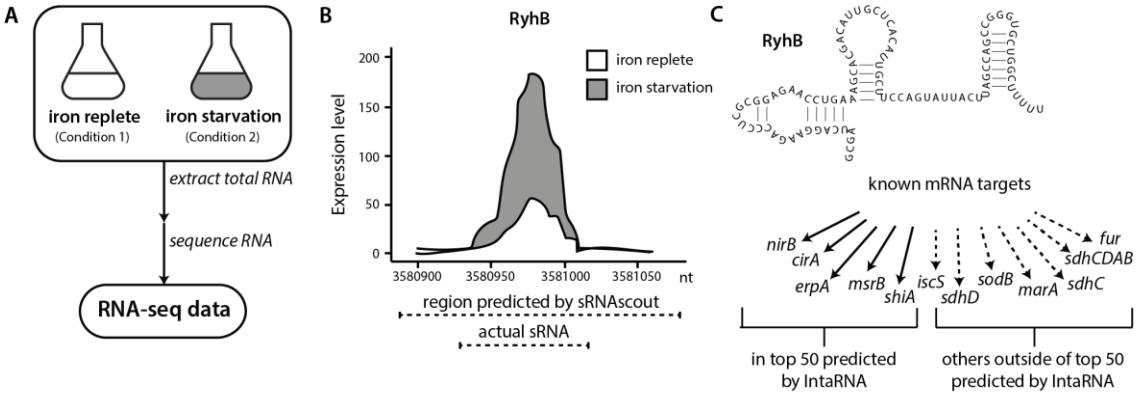


Figure 4.3: sRNAscout identifies well-characterized iron-starvation regulator RyhB from existing *E. coli* transcriptome data.

(A) Existing RNA-seq data in the NCBI SRA database includes that of *E. coli* K-12 grown in iron-replete and iron-starvation conditions [157]. **(B)** From this data, sRNAscout predicted sRNA RyhB based on its intergenic expression and differential expression between the two growth conditions. RyhB expression is higher in the iron-starvation than the iron-replete condition. The actual RyhB transcript is completely within the region predicted by sRNAscout with about 50 nt extra on each side. **(C)** RyhB has 12 mRNA targets with verified binding sites. Five of these were predicted by IntaRNA to be in the top 50 mRNA targets among all mRNAs in the transcriptome. sRNAphenoscore uses the top 50 IntaRNA targets to predict each sRNA candidate's phenotypic impact.

For this part of the study, we processed RNA-seq raw files according to the REFINE bioinformatics workflow (**Figure 4.2**). As shown in **Figure 4.3B**, RyhB shows ~5-fold higher expression upon iron starvation and its peak reaches above the minimal expression requirement of 63 counts. Of the 12 known mRNA targets of RyhB (**Figure 4.3C**), IntaRNA predicted five in the top 50 of favorable energy by IntaRNA: *nirB* (ranked

9), *cirA* (18), *erpA* (27), *msrB* (31), and *shiA* (38). We note that additional real targets of RyhB (not yet clarified with specific binding sites) may exist within the top 50 IntaRNA targets; some included in these 50 suggested by previous omics data are *nuoG*, *dmsA*, *dmsC*, *hypB*, *narG*, *grxD*, *sufB*, *cysE*, and *metH* [153]. The differential expression of each of the top IntaRNA-predicted targets contributes to the sRNA's phenoscore. Combining the contribution of RyhB differential expression (sRNAscore) and the differential expression of its predicted targets (mRNAscore), its total phenoscore is 0.658, ranking it 1652 among sRNA candidates predicted by sRNAscout.

Several known sRNAs in *E. coli* ranked highly with sRNAscout for iron-starvation stress response regulation including MgrR (rank 2), DsrA (25), RydC (63), and RyfA (64). These are each known to respond to or regulate stress responses other than iron-starvation. Overexpression or deletion of each will reveal its phenotypic impact under this stress.

Additionally, we uncovered at least one novel sRNA in *E. coli* that responds to iron starvation stress, possRNA340 (**Figure 4.4**). This transcript ranked 14 by phenoscore from sRNAscout, mainly due to its dramatic differential expression between iron-replete and iron-starvation condidtions. The northern blot of this sRNA shows two transcripts, one the length of the predicted region by sRNAscout (195 nt) and the other slightly larger (310 nt) (**Figure 4.4**). Further testing will reveal any regulatory effects of this sRNA.

Overall, the application of REFINE to this iron-starvation stress phenotype in *E. coli* demonstrates the ability to find a real, impactful sRNA and novel regions to be examined for future engineering.

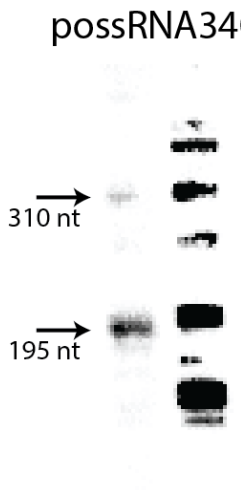


Figure 4.4: Novel sRNAs in *E. coli* discovered by sRNAscout.

E. coli sRNA candidates among the top 50 by phenoscore in iron-replete vs. iron-starvation conditions were tested by northern blot to verify detectable expression. In each blot, the first lane includes 30 µg total RNA of wild-type *E. coli* K-12 grown in iron-replete conditions and collected in stationary phase. The second lane includes ΦX174 DNA/HinfI Dephosphorylated Markers.

4.3.3 Identification of sRNAs in *Z. mobilis* that Improve Ethanol Tolerance

As a second case for this study, we aimed to engineer ethanol tolerance in *Zymomonas mobilis*, by identifying novel sRNAs (that, unlike RyhB, were previously uncharacterized). *Zymomonas mobilis* naturally produces ethanol up to 12% (v/v) ethanol and tolerates up to 16% (v/v), making it an especially interesting bacterium for biofuel applications [34,35,49]. Over the last 20 years, metabolic engineering and directed evolution developed a variety of *Z. mobilis* strains [59,60]. Transcriptome data are available for *Z. mobilis* under a variety of stresses including ethanol, furfural, acetate, and oxygen [48,50–53]. Although ethanol stress RNA-seq data is not available for *Z. mobilis* (only microarray transcriptomes), anaerobic and aerobic data has been collected. Anaerobic conditions facilitate higher levels of ethanol production relative to aerobic conditions [158,159] and

therefore are routinely used to represent conditions of lower and higher ethanol, respectively.

Recently discovery of 15 sRNAs in *Z. mobilis* included four responsive in expression to oxygen or ethanol stresses, representing potential regulators for engineering robustness [53]. However, this study a systematic approach to identify sRNAs specifically useful to enhance ethanol tolerance. Moreover, the low-throughput process of identifying these sRNAs included manual searches through transcriptome data, followed by northern blotting.

To identify ethanol-enhancing sRNAs in *Z. mobilis* in a high-throughput manner, we processed the raw RNA-seq files and used sRNAscout to predict sRNAs (**Figure 4.3A**). With the filtering criteria (**Table 4.1**), sRNAscout predicted 679 sRNA candidate regions (**Table A.9**). We adjusted the expression criteria from the *E. coli* iron-starvation values to be less stringent because this *Z. mobilis* data retains rRNA and therefore exhibited more shallow sequencing depth. Of the previously confirmed sRNAs, sRNAscout identified all as sRNA candidates except Zms2 (as a note, Zms13 and Zms14 (ENA ASM710v1) have now been annotated as tmRNA and CRISPR, respectively). **Figure 4.5** shows the expression profiles of these sRNAs and the specific regions predicted by sRNAscout. Note sRNAscout probably failed to detect Zms2 because it shows the lowest expression and least differential expression among these verified sRNAs. As shown in **Figure 4.5**, the sRNA candidate genomic coordinates predicted by sRNAscout generally correlate with sRNA transcript ends experimentally verified by RACE [53]. All of the sRNAscout-predicted regions fully contain the real transcript such that cloning of these regions without previous knowledge of the ends should produce functional transcripts suitable for screening for phenotype impacts.

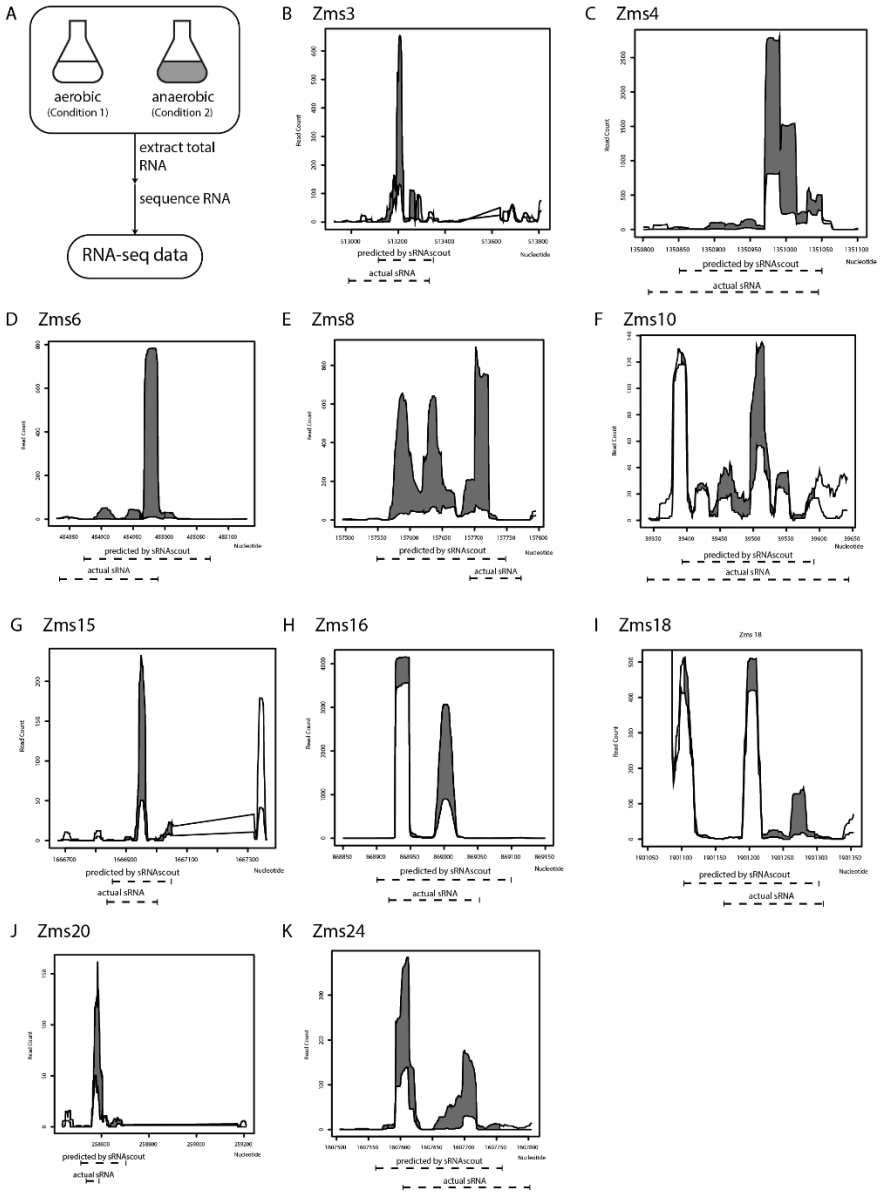


Figure 4.5: sRNAscout identifies previously known *Z. mobilis* sRNAs responsive to oxygen stress.

(A) In a previous study, *Z. mobilis* ZM4 was grown in aerobic and anaerobic conditions and the RNA was extracted and sequenced [53]. Original manual analysis of these data identified 13 sRNAs which were verified by northern blot. In this study, these data were analyzed with the REFINE approach to find the sRNAs automatically. **(B-L)** The expression profile of each known sRNA is shown for the aerobic and anaerobic conditions, as well as the predicted coordinates by sRNAscout compared to experimentally verified ends.

Table 4.1: Filtering criteria used by sRNAscout to identify sRNA candidate peak regions in each dataset.

Criteria	*Z. mobilis* aerobic vs. anaerobic	*E. coli* iron starvation
Minimum expression level	34 counts	63 counts
Consecutive nt with minimum expression	24 nt	20 nt
Differential expression level	2.7-fold	3.38-fold
Consecutive nt with differential expression	15 nt	8 nt

Table 4.2: Phenoscores from aerobic vs. anaerobic data of previously discovered *Z. mobilis* sRNAs.

These ranks correspond to the phenoscores across all 679 sRNA candidates predicted by sRNAscout in these conditions.

sRNA	mRNAscore	sRNAscore	phenoscore	rank
Zms6	0.580	0.781	1.361	16
Zms8	0.670	0.535	1.205	53
Zms20	0.785	0.406	1.191	70
Zms18	0.751	0.351	1.101	133
Zms3	0.705	0.167	0.873	347
Zms24	0.639	0.226	0.864	360
Zms4	0.719	0.101	0.820	427
Zms16	0.661	0.142	0.802	465
Zms10	0.639	0.125	0.764	520
Zms15	0.688	0.000	0.688	617

Using IntaRNA, we next predicted the targets for each sRNA candidate region and the top 50 predictions (lowest energy) were used to calculate the mRNAscore. As shown in **Table 4.2**, the phenoscores of previously verified sRNAs showed a range comparable to the overall range of all sRNA candidates predicted by sRNAscout, suggesting a variety of potential impacts on the ethanol tolerance phenotype. It is worth noting that the contribution of the sRNAscore (SD = 0.153) seems more indicative of the overall

phenoscore for this subset of verified sRNAs compared to the mRNAscore, which shows less variability (SD = 0.045).

Of the 50 sRNA candidates with highest phenoscores, eight were tested by northern blot in a previous study and only Zms6 was detected [53]. Thirty additional sRNAs were selected and tested for expression by northern blot. Six were confirmed to have detectable expression in stationary phase (**Figure 4.6**), representing novel transcripts that respond to ethanol stress.

To test the impact of these sRNAs on the ethanol-tolerance phenotype, we next cloned 6 sRNAs into plasmids for overexpression under the constitutive Pgap promoter (native to *Z. mobilis*, **Figure 4.7A**). Additionally, to observe the impacts of sRNAs with a range of phenoscores, we cloned the previously discovered sRNAs (ranging from 16 to 617 in the ranked phenoscore list).The strains were screened for their growth rates with and without 8% (v/v) ethanol supplemented to the media (**Figure 4.7B**). As shown in **Figure 4.7C**, the overexpression of Zms4 and Zms13 negatively impact growth rate in ethanol compared with the pEmpty control while Zms8, Zms9, and Zms16 overexpression strains show higher relative growth rates than pEmpty strain. Among the set of sRNAs screened here, Zms8 has both the highest phenoscore as well as a significant impact on growth in ethanol stress.

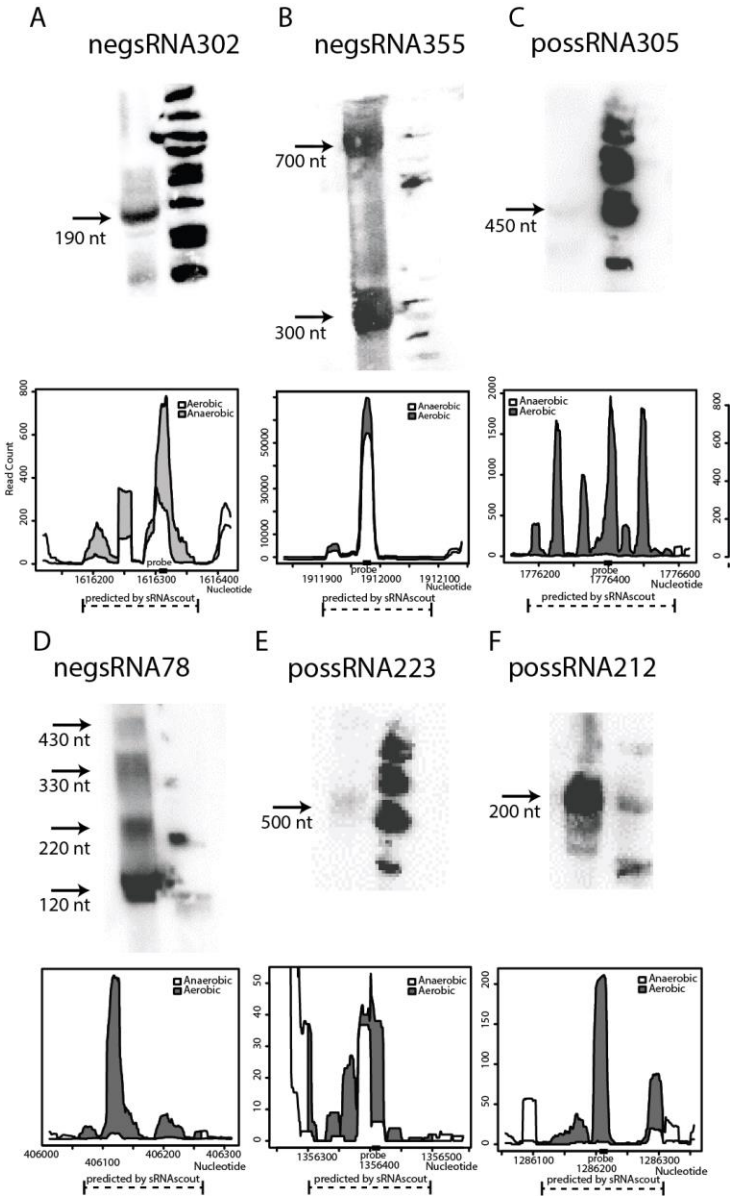


Figure 4.6: Novel sRNAs in *Z. mobilis* discovered by sRNAscout.

Z. mobilis sRNA candidates among the top 50 by phenoscore in aerobic vs. anaerobic conditions were tested by northern blot to verify detectable expression. In each blot, the first lane includes 30 µg total RNA of wild-type *Z. mobilis* 8b grown in anaerobic conditions and collected in stationary phase. The second lane includes ΦX174 DNA/HinfI Dephosphorylated Markers. Each confirmed sRNA is shown along with its expression profile according to the RNA-seq data. A dashed line marks the genomic region predicted as an sRNA candidate by sRNAscout and a solid rectangle marks the location of the complementary probe designed for northern blotting.

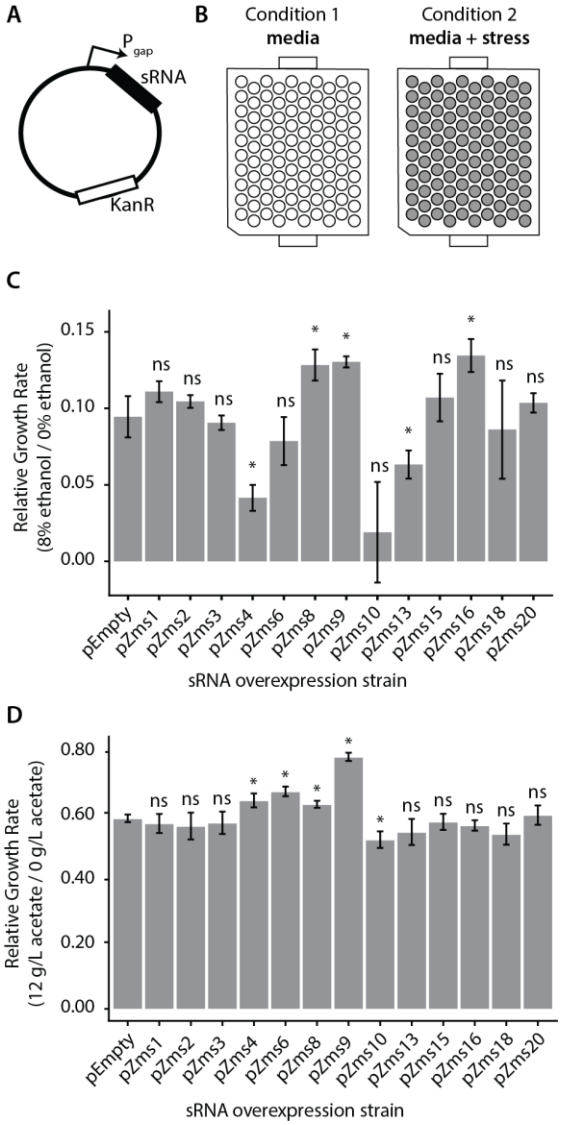


Figure 4.7: *Z. mobilis* sRNAs identified by the REFINE process impact stress tolerance.

(A) Each previously identified *Z. mobilis* sRNA was cloned into a plasmid under the constitutive P_{gap} promoter for overexpression. **(B)** *Z. mobilis* strains each carrying an sRNA overexpression plasmid were grown in 300 uL plate cultures in two conditions: media only and media with added stress conditions (8% v/v ethanol or 12 g/L acetate). Cell density (OD_{600}) was measured every 15 min for 24 h by the Bioscreen C and maximum growth rates were calculated. **(C-D)** Growth rates of each strain were normalized to their own growth rate in the no-stress condition (grey bars, error bars = SD, n = 3). The relative growth rate of each overexpression strain was compared by t-test to the pEmpty plasmid strain which lacks an sRNA (* $p < 0.05$). Multiple sRNAs show significant impact in 8% v/v ethanol stress **(C)** and in 12 g/L sodium acetate stress **(D)**.

To discern if these sRNAs respond specifically to ethanol or could be involved in other stress responses, we tested these overexpression strains were grown in acetate stress as well (**Figure 4.7D**). Along with the desired sugar monomers, acetic acid is released during the industrial pretreatment process of cellulose and is a strong growth inhibitor [160]. Overexpression of Zms4, Zms6, Zms8, and Zms9 increase growth rate while overexpression of Zms10 significantly decreased growth rate in acetate compared to the control strain. These results suggest that Zms4, Zms8, and Zms9 may play general stress response roles and be beneficial for a number of strain engineering goals.

4.4 DISCUSSION

The REFINE approach exploits transcriptome data to inform engineering complex phenotypes with sRNAs. The bioinformatics process includes a novel method, sRNAscout, for identifying sRNAs that impact the phenotype of interest without previous characterization. In this study, we identified novel sRNAs responsive to changing oxygen levels including negsRNA302. Additionally, the program sRNAphenoscore assigns each sRNA candidate predicted by sRNAscout a score representing its predicted impact on the target phenotype. In this way, we prioritize the list of predicted sRNA for follow-up characterization to those most likely useful for strain engineering.

The sRNAscout program was inspired by sRNA discovery efforts based on differential expression of novel transcripts between growth conditions. The sequence conservation-based sRNA prediction tool SIPHT previously predicted 143 sRNAs in *E. coli* K-12, including RyhB. Only 8 of the 143 SIPHT regions were identified by sRNAscout, illustrating again that the sequence-based and transcriptome-based approaches detect different kinds of sRNAs. Similarly, 2 of the 4 SIPHT regions predicted in *Z. mobilis*

were also found by sRNAscout. The sRNAscout tool offers targeted discovery of transcripts responsive to the growth conditions of interest.

As strain engineering moves increasingly to fully automated systems, an approach like REFINE is very attractive. With transcriptome data from two or more growth conditions, REFINE can computationally identify impactful genomic regions that can be targeted for mutations. Mutant strains can be cloned and screened in large libraries to find the most impactful mutations and then explore which combinations produce the fittest strain. Note that for high-throughput, automated strain development an expression-based approach like sRNAscout, which finds any/all impactful genomic regions for mutation and screening, is preferable to a purely sequence-based approach. This is because it can identify any intergenic transcripts expressed differentially under the stress conditions, including undiscovered transcripts that may not be sRNAs, whereas a sequence-based approach would exclude these other non-sRNA discoveries.

A challenge we foresee for full automation is the identification of transcript ends for cloning. As shown in **Figure 4.3B** and **Figure 4.5**, sRNAscout ends do not fully align with ends discerned by RACE in previous studies. sRNAscout identifies peaks of expression and, if less than 200 nt, the program expands the sRNA candidate region to 200 nt. For the transcripts to be functional as regulators, they must include the binding sites of their targets and exhibit sufficiently similar folding with the native transcripts. To identify the ends of sRNA candidates, the user may employ a computational approach like DeepBound [161] or a traditional experimental approach like 5' and 3' RACE. In some cases, it may be feasible to clone and test the entire intergenic region or to split it into multiple short fragments for testing. This will depend on the volume of cloning and screening that can be done and on the depth of sRNA characterization desired (and already existing for that organism). For *E. coli*, we expect that previously predicted or confirmed sRNAs will

appear in the sRNAscout predictions and these can inform transcript end definitions as well.

It is important to note that the sRNAs discovered by REFINE may or may not be direct regulators of the phenotype of interest. The bioinformatics approach is designed to predict the direct impacts because it scores sRNA candidates based on the direct binding of sRNA-mRNA pairs predicted by IntaRNA. But the differential expression of these mRNA targets could arise from direct sRNA regulation, indirect regulation, or no real regulation at all. It is impossible to decouple these things without more data and ultimately not necessary to identify impactful sRNAs. The native regulatory landscape is very complex. Some sRNAs may regulate multiple stresses while others play a very specific role. With the *Z. mobilis* sRNA overexpression strains screened here, we found that Zms13 and Zms16 showed impact specific to ethanol stress whereas Zms4, Zms8, and Zms9 also had an impact on acetate stress. Additionally, sRNAs may work together with other sRNAs or regulators to coordinate the complex phenotype and multiple. Ultimately, we aim to find sRNAs impactful to phenotype whether general regulators or specific to a single phenotype.

Depending on the organism and available genomic manipulation methods, sRNA candidate overexpression on a plasmid may be replaced by genomic overexpression, deletion, or knockdown. It is important to note that overexpression on a plasmid removes the transcript from its native regulatory context and may not give the full picture of an sRNA candidate's phenotypic impact. However, considering the desired application of strain engineering (rather than sRNA characterization), transcripts that can be successfully manipulated to achieve a new phenotype carry the most value, so those very sensitive to additional layers of regulation for their activity may not be desirable anyway.

The increasing availability of RNA-seq data and accessibility of high-performance computing resources continue to increase fuel great potential for wide application of the REFINE approach for strain engineering.

Chapter 5: Major Findings and Perspectives

In this work, I contribute multiple new strategies to identify and characterize ncRNA regulators useful for strain engineering. These strategies enabled us to discover *cis*-acting and *trans*-acting ncRNAs in *Z. mobilis* and to map their regulatory roles in stress tolerance. Overall, this work serves to accelerate complex phenotype engineering in a variety of organisms.

In the context of local, *cis*-acting RNA regulators, we developed a strategy to identify 5'UTRs from transcriptome data and to rapidly screen their responses to stress in vivo [162]. This approach is distinct from conventional methods that rely on sequence conservation with other organisms. Many of these sequence-based algorithms only yield previously known types of UTRs in organisms closely related to the model organisms on which the algorithms were trained. GC-rich organisms present a particular challenge to these algorithms and real genes and regulatory elements go undetected. An advantage of our transcriptome-based approach is that it can be targeted to identify useful regulatory elements related to growth conditions (and metabolic phenotypes) of interest. Additionally, it unlocks discovery of new, unique 5'UTRs missed by previously developed approaches. Coupled with the in vivo fluorescence-based screening system developed in this work, we highlight the feasibility of coupling the use of transcriptome data, bioinformatics analysis, and fluorescence-based screening to identify novel regulatory 5′UTRs in other microorganisms.

With the 5'UTR discovery pipeline, we exposed the first 5'UTRs in *Z. mobilis* and illuminated their roles in stress responses. A total of 36 novel 5'UTRs emerged from this study. Two of these influence expression of their mRNA targets in response to ethanol stress: UTR_ZMO0347 (RNA binding protein Hfq) and UTR_ZMO1142 (thioredoxin

reductase). UTR_ZMO1142 also responded to xylose and acetate stresses, suggesting it may be a general stress response regulator. A multi-stress responsive regulatory element like this could be attractive for engineering robustness to multiple stresses, as demanded by the industrial setting.

Considering the important role of Hfq as a protein chaperone for *trans*-acting ncRNA control, it is fascinating that it is itself controlled by a 5'UTR. This UTR does not show sequence conservation among any other bacterial species. Previously, *Z. mobilis* Hfq showed involvement in conferring tolerance to multiple industrial inhibitors like acetate, vanillin, furfural, and HMF [48,49,83]. Similar roles of Hfq in pathogen stress responses inspired the strategy of Hfq deletion to hinder the ability to adapt to the stress of their hosts [163–165]. However, in cases where we aim to improve strain tolerance with Hfq, specific strategies are less clear. Our discovery of the Hfq 5'UTR response to ethanol motivates further investigations of Hfq and how it might be manipulated for engineering alcohol tolerance in *Z. mobilis* (and perhaps other bacteria). This example of multi-layer regulation points to the enormous challenge of continuing to map these networks to exert proper engineering control.

In general, 5'UTRs represent useful biological parts that can be readily transferred to other genes or to other organisms due to their small size and direct response to environmental stimuli. Further investigation of natural 5'UTR mechanisms may inspire new synthetic designs or design rules.

Strategies for *trans*-acting ncRNA discovery and characterization were also presented in this work. *Trans*-acting ncRNAs naturally exert control on multiple genes and pathways, making them especially powerful candidates for engineering complex phenotypes. However, their impact has remained largely untapped in metabolic engineering because of their lack of foreseeable functions (as their characterization has

been slowed relative to their discovery). Here, we contributed two major steps forward: (1) we characterized a network of ncRNA regulators related to ethanol tolerance in *Z. mobilis* and (2) we developed a general bioinformatics pipeline (REFINE) to mine transcriptome data and produce a list of most likely ncRNAs relevant to a target phenotype.

We uncovered a network of *trans*-acting ncRNAs involved in the natural regulation of ethanol tolerance in the efficient ethanol producer *Z. mobilis*. Through integrated omics and biochemical methods, we identified and characterized ncRNAs Zms4 and Zms6 along with their interacting transcript targets. These are the first ncRNA-mRNA pairings known in this organism and the first identified ncRNA network with relevance to ethanol stress in bacteria.

A unique feature of this ncRNA network is that multiple ncRNAs physically bind to each other as well as to multiple mRNA targets. To our knowledge, this is the largest known network of directly interacting ncRNAs that co-regulate a number of targets to confer a specific complex phenotype, introducing new questions about how ncRNAs cooperate to manage network responses. We expect more examples of mutli-ncRNA networks to emerge in the coming years and the synergy between these regulators to be key in engineering efforts.

In the absence of Zms4 or Zms6 (through genetic knockout/knockdown), *Z. mobilis* is highly sensitive to ethanol stress. When we altered expression levels of these ncRNAs, we observed significant changes in ethanol tolerance and production. We additionally showed a number of integrated pathways including transport, oxidoreductase activity, and lipid metabolism that are co-regulated to confer tolerance to ethanol (as indicated by the fact that mRNA targets of these ncRNAs participate in these pathways). Similar ncRNAs and integrated pathways likely exist in other organisms that exhibit high alcohol tolerance.

As we began to investigate this ncRNA network in *Z. mobilis* and spend time in reviewing literature, we recognized ncRNAs as ideal resources for regulating multiple metabolic pathways that make up complex phenotypes. A major barrier preventing their use was the gap between discovery and mechanistic analysis. Individual functional screening and characterization of large numbers of potential ncRNA regulators identified by sequence or transcriptome approaches was highly inefficient, as many of those characterized exhibited no relevance to desired phenotypes. We realized the need for an fast and generalizable method for metabolic engineers to find naturally occurring ncRNA regulators relevant to desired phenotypes.

The computational tools in the REFINE pipeline (sRNAscout and sRNAphenoscore) aim to end to manual scrolling through transcriptome data to find ncRNAs that are truly expressed and have a desirable functional impact. As the availability of RNA-seq data and accessibility of high-performance computing resources continue to increase, we anticipate that REFINE will become increasingly useful to analyze the transcriptome of a variety of microbes to achieve a variety of complex phenotypes.

As strain engineering moves increasingly to fully automated systems, an approach like REFINE is very attractive. With transcriptome data from two or more growth conditions, impactful genomic regions can be computationally identified and targeted for mutations. There is great potential for the REFINE approach to be widely applied for complex phenotype engineering efforts.

The pipeline could be adapted for discovery of other types of regulators beyond *trans*-acting ncRNAs. For example, during its development, we noticed that the sRNAscout prediction parameters could be adjusted to uncover 5' and 3' UTRs or to detect "peaks" of expression in mRNA regions. The REFINE approach could be extended to include proteomics and metabolomics for even more powerful prediction of ncRNA

regulators. Also, we could integrate recent developments in our lab to profile RNA structural landscapes in vivo with high-throughput sequencing with the REFINE pipeline to enhance the ncRNA-mRNA interaction scores with experimental data (in addition to the IntaRNA prediction approach employed thus far) [166].

In *Z. mobilis*, several challenges remain for its development into an advanced industrial producer of chemical products from biomass. Primarily, tolerance to multiple simultaneous stresses needs to be improved. Our work presents major steps forward by expanding the regulatory parts and computational tools available to achieve complex phenotype goals with noncoding RNAs.

More broadly, this work draws out some key threads in the field of biological engineering. First, big data does not directly translate to big impact. Enormous collections of genomes, transcriptomes, proteomes, and metabolomes contain a wealth of insight, but their impact on strain engineering has been more modest than desired. We need analytics to harvest and refine the data into executable strategies. Second, ncRNAs belong in the metabolic engineering toolbox. These dynamic regulators exist in nature to manage complex phenotype responses and their impact still remains largely unexplored, especially in nonmodel organisms. Finally, many industrially relevant phenotypes are highly complex and we continue to uncover additional layers of regulation that reroute metabolism (especially during metabolically-induced stresses). For instance, in this work alone, we observed ncRNAs interacting with ncRNAs and a 5'UTR regulating a protein regulator of ncRNA regulators. Although these layers can be confounding, they are also inspiring. The potential to combine multiple types of regulation into cohesive strategies will keep us innovating for years to come.

Chapter 6: Materials and Methods

6.1 Methods for Chapter 2

6.1.1 Bacterial Strain and Culture Conditions

Z. mobilis 8b was provided by NREL [60]. Cells were cultured in 5 mL RMG (glucose, 20.0 g/L; yeast extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) [48] overnight at 33°C and then inoculated into 100 mL fresh medium. Initial OD_{600nm} was ~0.05 and cells were induced with 10 μg/mL of tetracycline when OD_{600nm} reached ~0.4. 1, 3, 5, or 8% (v/v) ethanol was supplemented into RMG for ethanol stress experiments from the beginning of growth. 10 g/L of sodium acetate (NaAc) [51] was supplemented into RMG for acetate stress. In the case of xylose stress, 10 g/L glucose + 10 g/L xylose was used instead of 20 g/L glucose in the media. Cells were collected 4 or 12 h post-induction for measuring GFP expression.*

6.1.2 Construction of *hfq* Mutant Strains

Upstream and downstream fragments (each about 1 kb) homologous to the target deletion gene were assembled with the spectinomycin gene *aadA* in the middle. The upstream homologous arm was PCR amplified using primers Fup and Rup, and the downstream homologous arm amplified by Fhfq-1, Fhfq-2, and Rhfq. Primers Fsp, Rsp-1, and Rsp-2 were used to introduce a terminator at the end of the spectinomycin resistance gene and primers Fhfq-1 and Fhfq-2 were used to introduce the 3xFLAG. The fragments were then assembled similarly to Gibson assembly. Briefly, primers were designed with 15-nt overhangs of desired upstream or downstream sequences and used to amplify the pieces for the assembly and purified from the primer dimers by gel electrophoresis

* Section 7.1 has been previously published.[162]

followed by gel purification. Fragments and the pUC57 vector were mixed in a molar ratio of 3:1 (about 120 ng total DNA) and added to 0.5 U T5 exonuclease (NEB, USA), 0.5 µL Buffer 4 (NEB, USA), and nuclease-free water (up to 5 µL). The reaction mix was kept on ice for 5 min and then transformed into *E. coli* competent cells by chemical transformation. Cells were plated on LB agar plates with 200 µg/mL of spectinomycin. Transformants were screened by colony PCR and confirmed by Sanger sequencing (Sangon, China). The purified plasmid was electroporated into the *Z. mobilis*. Transformants appearing on RM agar plates with 200 µg/mL of spectinomycin were cultured and screened by colony PCR. Colonies with correct PCR product sizes were selected as deletion candidates after Sanger sequencing confirmation.

6.1.3 5' Rapid Amplification of cDNA Ends (RACE)

RACE experiments were performed on total RNA samples using FirstChoice® RLM-RACE kit (Ambion, CA) according to the manufacturer's protocol and to previously published work [53]. Briefly, 8 µg RNA was treated with Tobacco Acid Pyrophospatase (TAP) at 37°C for 1 h, followed by ligation of the 5' RACE kit adapter at 37°C for 1 h. The resulting RNA was then reverse transcribed according to the manufacturer's protocol and PCR was performed on the resulting cDNA. All primer sequences used for RACE are listed in **Table A.3**. Resulting PCR products were purified using QIAquick PCR purification kit (Qiagen, MD) and RNase-free water (Ambion, CA) for final elution. Final PCR products were sequenced and results were compared with the genome. 5' RACE adapter sequences (5'-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA-3') and adjacent mRNA sequences were used for the detection of TSS in the 5'UTRs.

6.1.4 Construction of GFP-Reporter Plasmids with 5'UTRs

The tetracycline-inducible promoter with GFP construct (pEZ-tet-GFP) was derived from the minimal shuttle vector pEZ15Asp of *Z. mobilis* and contains the replication origin of *E. coli* and *Z. mobilis* [90]. A parental plasmid containing *pheS* counter-selection marker [167,168] was constructed with *pheS* counter-selection marker incorporated in front of the *gfp*gene flanked by BsmBI sites (Type IIS enzyme), which is one of the Type IIS restriction endonuclease to enable Golden Gate cloning for efficient cloning of each 5'UTR-containing GFP [169] (**Figure 2.6E**). With this design, 5'UTRs along with the first 90-bps of the downstream mRNA were cloned for each candidate right in front of the *gfp* gene in frame. This design was implemented to preserve the native structure of the UTR by better mimicking its native context while still allowing GFP expression. Primers used for the amplification of UTR+90-bps are listed in **Table A.3**. Each primer contains a BsmBI enzyme site on the 5' end.

6.1.5 Fluorescence Measurements

Cells were analyzed by flow cytometry using the FACSCalibur™ (BD Biosciences, CA) according to standard methods as described in a previous study [168]. Briefly, cells were pelleted and resuspended in phosphate buffered saline (PBS: 137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM phosphate buffer, pH 7.5) to a concentration on the order of 10^7 cells/mL. The cells were excited with the 488 nm argon laser and the cell population was determined from the forward scatter and side scatter distributions reported by the cytometer. Data were collected for at least 50,000 active cells, ensuring enough events to assume that the population distribution would be unaffected by rare events. Sample data were analyzed using CellQuest Pro (BD Biosciences) with a user-defined gate.

Averages of median values for each sample were calculated from biological triplicates and error bars were calculated as standard error of the mean.

6.1.6 Quantitative RT-PCR

RT-PCR was used to experimentally confirm the presence of 5'UTRs from the candidate list (**Table A.1**). Total RNA was prepared using Trizol reagent (Invitrogen, CA) and resulting RNA was treated with DNase I (RNase free, ThermoScientific, MA) to prevent of genomic DNA contamination as described in the manufacturer's protocol. Subsequently, RNA was precipitated with isopropanol and then washed with ethanol. Total RNA rehydrated with nuclease-free water (Ambion, CA) was used as the template for reverse transcription. Two hundred nanograms of RNA was incubated for the first strand synthesis with 100 ng of random hexamer and 10 mM dNTPs at 65°C for 5 min. According to the manufacturer's protocol, SuperScript™ III Reverse Transcriptase (Invitrogen, CA) was added to RNA-primer mix with RNaseOUT™ RNase Inhibitor, 0.1 M DTT, 5x First-strand buffer and then incubated for 5 min at room temperature. The final reaction mixture was incubated at 55°C for 1 h and then heat inactivated at 70°C for 15 min.

The cDNA product from first-strand synthesis was used as a template in a 20 µL PCR reaction containing Power SYBR® Green PCR master mix (Invitrogen, CA). Specific primers were used for each target. For example, primers F (5'-GCGTCTTGTTGACCCGTAAT−3') and R (5'- AATCCTCGTCTCGCCTTTCT−3') were used for the amplification of the *hfq* gene. Three primers were designed for 5'UTR RT-PCR: the first forward primer (with reverse primer, set A) was located in the middle of the 5'UTRs and the second forward primer (with reverse primer, set B) was designed to bind in the front part of following mRNA regions. The reverse primer was designed to bind in the middle of the mRNA regions. All primers used in this study are listed in **Table A.3**.

Phusion® High-Fidelity DNA Polymerase (NEB, MA) was used for PCR amplification. No reverse transcriptase was used for the negative control to exclude potential genomic DNA contamination. Primer set B was used for the positive control as it represents amplification of the mRNA coding region. The temperature cycle used for the PCR reactions is as follows: 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 60 s. Relative quantification of *hfq* transcript between each condition was performed using Viia 7 Software (Life Technologies, CA) following the comparative delta-delta threshold cycle ($\Delta\Delta C_T$) method. Samples were collected in biological triplicates.

6.1.7 Western Blotting Analysis and Quantification of Protein Expression Levels

Western blotting analysis was performed to detect GFP expression using anti-GFP antibody (Roche 11814460001). Standard western blotting protocols were used as previously described [170]. Briefly, total cellular lysates were loaded onto a 12% denaturing SDS-PAGE gel. Gels were transferred to methanol activated PVDF membranes using the Trans-Blot® Semi-Dry Electrophoretic Transfer Cell (Bio-Rad, CA) and run for 40 min at 15 V or 20 min at 25 V. Membranes were blocked with 5% dry non-fat milk in Tris-buffered saline (TBS, pH 7.5) for 1 h at room temperature. The proteins were detected with anti-GFP antibody at 1:1000 dilutions or anti-FLAG at 1:5000 (Proteintech, China). As a secondary antibody, goat anti-mouse IgG (H + L) HRP Conjugate (Promega #W4021) was used at a dilution of 1:2500 (or 1:5000 with Proteintech, China). All images were developed using Clarity™ ECL Western Blotting Substrate (BioRad, #170-5060) and the ChemiDoc™MP Imaging System (BioRad, CA) or West Dure Extended Duration Substrate Kit (AntGene, China) with AI600 Imaging System (GE, USA). Bradford assay measurements were used to normalize the loading of all protein analyses by total protein mass. Specific proteins were detected on the membrane by western blot analysis and

quantified using ImageQuant TL 8.1 (GE Healthcare, CA) or Gel-Pro analyzer (LiuYi, China). Each protein was detected using anti-GFP. The level of GFP expression was measured and then normalized using the expression of RecA as an internal control.

6.2 METHODS FOR CHAPTER 3

6.2.1 Strains and Culture Conditions

The *Z. mobilis* 8b strain, which is an integrant of the ZM4 strain (ATCC 31921), was used in this study. *Z. mobilis* 8b was cultured in RMG media (glucose, 20.0 g/L; yeast extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) at 33°C. *E. coli* DH5α was used for plasmid construction and manipulation, grown in LB media at 37°C.

To generate sRNA overexpression strains, each sRNA was cloned into the NcoI-SalI site of the pEZ-tet vector [90] to allow for inducible expression under the P_{tet} promoter. For sRNA co-immunoprecipitation constructs, a gBlock® (NEB) each of 2MS2BD-Zms4/Zm6/control was cloned into pBBR1MCS2-P_{gap} vector [171] between NheI and SalI. All sequences of primers used in this study are listed in **Table A.6**. Strains containing pBBR1MCS2-P_{gap} plasmids were cultured with 350 μg/mL of kanamycin for *Z. mobilis* and with 50 μg/mL for *E. coli*. Overexpression strains containing pEZ-tet vectors were grown with 200 μg/mL spectinomycin for *Z. mobilis* and 50 μg/mL for *E. coli.*

For deletion strain construction, homologous upstream and downstream fragments (each 1 kb) of the target deletion gene (Zms4 and Zms6) were assembled with the spectinomycin gene *aadA*, flanked by LoxP sites, in the middle. Purified PCR product (1 μg) was directly electroporated (200 Ω, 25 μF, 1.6 kV) into *Z. mobilis.* Electroporated cells were recovered for 6 h and plated onto spectinomycin-containing plates (200 μg/mL). Plated cells were incubated anaerobically in a BD GasPak™ container system for 3-4 days

at 33°C. Transformants appearing on the selective plates were cultured and screened for correct size using PCR then sequence-verified.

Growth rates of mutant strains were evaluated using the Bioscreen C (Growth Curves US, NJ). Biological triplicates of each strain were grown in 5 mL RMG seed cultures with appropriate antibiotics at 33°C for 48 h. Cells were distributed in technical triplicates into Bioscreen C plates with RMG (with and without 8% ethanol) such that each well had a total volume of 300 µL and initial OD_{600} of 0.05. The Bioscreen C measured the turbidity with the wideband filter (420-580 nm) every 15 min for 48 h as the cultures grew without shaking at 33°C. The Bioscreen C was operated using EZ Experiment (Norden Logic Oy, Helsinki, Finland) and exponential growth rates were calculated using a custom MATLAB script [172].

For sRNA induction experiments, each strain was initially grown in biological duplicates in 5 mL RMG culture overnight then transferred into 500 mL with initial OD_{600nm} of 0.1. Cells were grown anaerobically at 33°C for 4 h to reach OD_{600nm} around 0.4. Then, 150 mL of cells were collected for proteomics, transcriptomics and ethanol assay. When OD_{600nm} reached 0.5 (~4.5 h of growth), 10 µg/mL anhydrotetracycline was added to each strain to induce sRNA expression from the plasmids. When OD_{600nm} reached around 0.6 (~5 h of growth), 150 mL of cells were collected to compare the gene expression profile in the middle of exponential phase. In this way, the effect of overexpressing Zms4 and Zms6 on the transcriptome and proteome could be confirmed by comparing the samples before and after induction. Final samples were collected during stationary phase (~12 h of growth). Pelleted cells were stored at -80 °C before further processing.

6.2.2 Ethanol Assay

Ethanol concentrations were measured using the UV-based ethanol assay kit (R-Biopharm, Darmstadt, Germany) according to the manufacturer's protocol.

6.2.3 RNA Purification

Cell pellets resuspended in 1 mL TRIzol reagent (Invitrogen) and transferred to screw cap tubes containing glass beads (Sigma) and incubated at 25°C for 5 min. Cells were lysed using a mini-beadbeater (BIOSPEC), with 100 s pulses three times. Cells were kept on ice for 10 min between each 100 s treatment. The beads and cellular debris were centrifuged at 4°C for 2 min. The supernatant was transferred to a clean siliconized 2 mL tube. After addition of 300 µL of chloroform:isoamyl alcohol mix (v/v 24:1), the samples were inverted for 15 s, and then incubated at 25°C for 3 min. Then, tubes were centrifuged at 13,000 rpm for 10 min and the aqueous top phase transferred to a clean siliconized 1.5 mL tube. Following this step, 270 µL of isopropanol and 270 µL of a mixture of 0.8 M sodium citrate and 1.2 M sodium chloride was added. The samples were mixed well, and then incubated on ice for 10 min. The RNA was pelleted by centrifugation at 13,000 rpm for 15 min. The pellet was washed with 1 mL 95% cold ethanol and centrifuged for 5 min. The pelleted RNA was allowed to air-dry for 5 min, and was resuspended in 50 µL RNase-free water (Ambion). RNA was digested with DNase I (RNase-free, ThermoScientific) for 1 h at 37°C to prevent genomic DNA contamination. DNase I was heat-inactivated by adding 0.5 mM EDTA to the reaction mixture and incubating at 75°C for 10 min. Then, RNA was incubated with isopropanol and GlycoBlue™ (ThermoScientific) at -20°C overnight. After centrifugation, pelleted RNA was washed with 95% cold ethanol and centrifuged. RNA was resuspended in 50 µL RNase-free water (Ambion) and stored at -80°C before sequencing.

6.2.4 Protein Purification

50 mL of cells were pelleted and resuspended in 5 mL phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Cells were lysed on ice by sonication (Qsonica 55 probe) at 10 W for 10 s followed by a 5 s incubation, a total of 30 times. The lysates were centrifuged at 4000 rpm for 10 min at 4°C and the supernatants containing soluble proteins were retained. Protein was precipitated by adding 4 volumes of ice-cold acetone and incubating at -20°C overnight. After centrifugation to pellet precipitated protein, the acetone was removed and the pellet air-dried for 30 min at room temperature. The protein pellets were resuspended in 100 µL of denaturing sample loading buffer and loaded onto an SDS-PAGE mini-gel (5% stacking, 10% resolving) and moved ~3 mm into the resolving gel by electrophoresis at 70 V. The gel was coomassie stained and total protein bands were excised and stored in destaining solution at 4°C.

6.2.5 RNA Sequencing and Analysis

Prepared RNA was quantified and checked for quality using Bioanalyzer before sequencing. NEBNext® Multiplex RNA Library Prep Set for Illumina® (New England Biolabs Inc.) was used for generating RNA libraries. Sequencing was performed using Illumina® NextSeq 500 with paired-end 2 x 150 bp runs (Genomic Sequencing and Analysis Facility at the University of Texas at Austin).

Adapter sequences and low-quality ends (phred quality < 30) were trimmed from the raw fastq files with cutadapt (v1.3) and reads shorter than 22 nt were discarded [173]. FastQC was used to verify good read quality for the trimmed files [174]. Reads were aligned to the *Z. mobilis* ZM4 reference genome (taxonomy ID 264203) with BWA-mem (v0.7.12-r1039) [175]. Aligned reads were filtered for quality (MAPQ ≥ 10) and sorted by

chromosomal coordinates using SAMtools (v0.1.19-44428cd) [176]. The number of reads aligned to each gene was calculated using HTSeq (intersection-nonempty mode for overlaps) [177]. DESeq2 was used to normalize and identify significantly differentially expressed transcripts between strains over time (likelihood ratio test of model ~strain+time+strain:time vs ~time; $p_{adj} < 0.01$) [178]. Data were visualized using ggplot2 [179] and Cytoscape Enrichment Map plugin [180,181].

6.2.6 Mass Spectrometry

Polyacrylamide gel bands containing protein were digested with trypsin. To identify proteins, LC-MS/MS was performed using the Dionex Ultimate 3000 RSLCnano LC coupled to the Thermo Orbitrap Elite with a 2 h run time at the ICMB Proteomics Facility using published methods [182]. Proteins were searched against the Uniprot *Z. mobilis* ATCC ZM4 database (April 27, 2016) using Sequest HT in Proteome Discoverer 1.4. The identifications were validated with Scaffold v4.4.1 (Proteome Software) with greater than 99.0% probability and with a minimum of two peptides at 99.0% peptide probability. In Scaffold, peptide and protein false discovery rates were both calculated as 0.0%.

6.2.7 MS2-affinity Purification of sRNAs in vivo

For use as an affinity tag, MS2 coat protein fused with maltose binding protein (MS2-MBP) [116] was expressed in *E. coli*. 100 mL of cells were cultured and induced with 1 mM IPTG at OD_{600nm} 0.5. Four hours after induction, cells were pelleted and resuspended in 10 mL column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM β–mercaptoethanol pH 7.4) and 2 mM PMSF (phenyl methylsulfonyl fluoride) to prevent degradation. After sonication on ice, lysates were treated with DNase I for 1 h at 4°C then centrifuged at 15,000 rpm. Supernatants were collected and suspended with amylose

magnetic beads (NEB). The MS2-MBP lysates were split into 200 µL aliquots, each washed with 1 mL column buffer twice. Lysates were incubated with washed amylose magnetic beads for 2-3 h at 4°C. Then, a magnet was applied and supernatants were decanted. Beads were washed with 1 mL wash buffer (column buffer + 0.1 mM maltose) three times. 50 µL of elution buffer (column buffer + 10 mM maltose) was added to the beads for the elution of MS2-MBP and incubated for 15 min at 4°C. To increase the yield, the elution step was repeated with 50 µL of elution buffer. Purified proteins were confirmed by SDS-PAGE gel and the concentration was measured using Bradford assay.

To find transcripts associated with each sRNA in vivo, total RNA (100 µL at 500 ng/µL) extracted from *Z. mobilis* strains containing pMS2, pMS2-Zms4, and pMS2-Zms6 was incubated with 2 µg purified MS2-MBP protein for 1 h at 4°C. Washed amylose beads were incubated with 2MS2BD-Zms4/Zms6/control+MS2-MBP complexes for 2 h at 4°C. Supernatants were removed from the beads by applying a magnet. Beads were washed three times with wash buffer and incubated with 50 µL of elution buffer for 15 min. The elution step was repeated so that 100 µL of each were collected. To precipitate the RNA, equal volumes of isopropanol and 10 µL of GlycoBlue™ was added to elution sample and incubated overnight at -20°C. RNA was pelleted at 15,000 rpm for 15 min at 4°C and washed with 1 mL ethanol. The RNA pellets were air-dried then resuspended in 50 µL RNase-free water. RNA samples were stored at -80°C until sequencing.

To find proteins associated with each sRNA in vivo, total lysates from *Z. mobilis* strains containing pMS2, pMS2-Zms4, and pMS2-Zms6 were incubated with 2 µg of purified MS2-MBP protein for 1 h at 4°C. Then, 1 mL Trizol was added to eluted samples for protein purification. 1.5 mL of isopropanol was added to the phenol-chloroform layer and mixtures were incubated for 10 min at room temperature and then centrifuged at 12,000 g for 10 min at 4°C to pellet the protein. Pelleted protein was washed with 2 mL of 0.3 M

guanidine hydrochloride in 95% ethanol and incubated for 20 min at room temperature then centrifuged at 7500 g for 5 min at 4°C. Washing steps were repeated twice more. Then, 2 mL of 100% ethanol was added to the protein pellets and samples were centrifuged at 7500 g for 5 min at 4°C. Air-dried protein pellets were resuspended in 3x SDS-loading buffer and run 3 mm into the resolving layer of an SDS-PAGE gel (5% stacking; 10% resolving) to concentrate protein before mass spectrometry.

6.2.8 PCR Amplification and in vitro Transcription of RNAs for Binding Assays

Primers were designed to amplify the 5' region between 200 nt upstream and 100 nt downstream of the transcription start site (ATG) of each mRNA target. If an annotated gene localizes upstream the targeted gene overlapping with the 5' region on the same strand, the forward primer was designed to anneal after the end of the upstream gene. For sRNAs, primers amplified the full length of the sRNA following the previously reported genomic coordinates [53]. In the forward primer, 4 random nt with the sequence "TACG" were introduce upstream of the T7 promoter sequence (TAATACGACTCACTATAGGGAGA) to facilitate the binding of the T7 polymerase. Following amplification, 5 µL of the PCR product was loaded on a 12% agarose gel to verify the size of the amplicon and the presence of a single product. Then, the remaining reaction volume was purified using the DNA Clean & Concentrator-5 kit (Zymo Research).

The in vitro transcription was conducted using the MEGA Script T7 kit (Thermo-Fisher Scientific) following the manufacturer instructions for up to 16 h of incubation at 37°C for transcription products shorter than 500 nt. Following TURBO DNase digestion, the RNA was separated by addition of 115 µL of water, 15 µl of Ammonium Acetate Stop Solution (included with the MEGA Script T7 kit) and 2 volumes of room temperature 25:24:1 Phenol: Chloroform Isoamyl Alcohol (PCA, Fisher BioReagents). This mixture

was inverted for 1 min, and then centrifuged at 15,000 x g for 2 min. The top aqueous phase containing the RNA was transferred to a new low-binding Eppendorf tube. Next, 1 µL of GlycoBlue (Thermo-Fisher Scientific) and 2.5 volumes of cold absolute ethanol (OmniPur, MilliporeSigma) were added. Then, samples were mixed well and incubated at -20°C overnight. The RNA was pelleted by centrifugation at 15,000 x g for 15 min, and washed with 500 µL of 70% cold ethanol. After centrifugation at 15,000 x g for 5 min, the pellet was dried under vacuum for 4 min using an Eppendorf™ Vacufuge™ Concentrator. Next, the pellet was resuspended in 10 µL of nuclease-free water and concentration was determined by spectrophotometry (NanoDrop 2000, Thermo-Fisher Scientific). To validate transcription, 500 ng of RNA was mixed with 2x RNA loading dye (NEB) and incubated for 5 min at 70°C. Next, samples were loaded in an 8% UREA gel (SequaGel, National Diagnostics) and run at 120 V for 2.5 h. After this step, the gel was stained with SYBR Green II (Thermo-Fisher Scientific) for 30 min, and visualized in a gel imager (ChemiDoc XRS$^+$ System, BioRad).

The internally labeled transcripts were synthesized as described above with the following modifications. The in vitro transcription reaction was scaled up to a reaction volume of 25 µL, and 2.5 µL of UTP [α-^{32}P] (3000 Ci/mmol 10 mCi/mL, 250 µCi, PerkinElmer) was added. Following ethanol precipitation and resuspension of the pellet in nuclease-free water, the unincorporated UTP nt were removed using DTR gel filtration cartridges (EdgeBio) as described by the manufacturer.

6.2.9 Detection of RNA-RNA Interactions

Each RNA of interest was amplified from *Z. mobilis* genomic DNA and in vitro transcribed as described in the Supplementary Material and Methods. The RNA binding reaction was performed in a 12 µL reaction volume containing 1x binding reaction buffer

(20 mM Tris–HCl (pH 8.0), 1 mM $MgCl_2$, 20 mM KCl, 10 mM Na_2HPO_4 -NaH_2PO_4 (pH 8.0)), 10% glycerol, 5 pmol of ^{32}P-labelled sRNA and 100 pmol of target RNA. This reaction was denatured at 70°C for 5 min, and then incubated at 37°C for 45 min. Samples were analyzed by electrophoresis in 5% glycerol and 5% nondenaturing polyacrylamide gel with 0.5x TBE running buffer at 4°C for 2.5 h. Next, the gel was placed on a sheet of Whatman grade GB004 blotting paper and dried at 80°C for 60 min (Gel Dryer 583, BioRad). Radioactive bands were visualized using a Typhoon FLA 700 (GE Health Life Science) and analyzed using CLIQS (TotalLab).

6.2.10 Northern Blot Analysis

Northern blotting analysis was performed by standard methods [53]. Briefly, DNA oligonucleotide probes specific for each sRNA were labeled using 20 pmol of oligonucleotide in a 20 µL kinase reaction containing 25 µM γ-P^{32} ATP and 20 units T4 polynucleotide kinase (NEB) at 37°C for 1 h. Ladder (ΦX174 DNA/*Hinf I* (Promega)) was labeled in the same manner. Total RNA (50-100 µg) from deletion and wild type strains were separated on a 10% denaturing polyacrylamide gel that was then was transferred to a positively charged membrane (Hybond N+, GE Life Sciences) for blotting. Hybridization was performed using Amersham Rapid-hyb buffer (GE Healthcare), following their recommended protocol for oligonucleotide probes, with overnight incubation at 42°C. After three washing steps with washing buffer (5x SSC, 0.1% SDS for the first washing and 1x SSC, 0.1% SDS for the second and third washing steps), membranes were exposed to a phosphor screen overnight and visualized with a phosphorimager (Typhoon Imager, Amersham Biosciences).

6.3 METHODS FOR CHAPTER 4

6.3.1 Calculating Intergenic Region Coverage from Existing RNA-seq Data

RNA-seq data from previous studies were analyzed as shown in **Figure 4.2A**. Raw fastq files were downloaded from NCBI SRA (*Z. mobilis* aerobic and anaerobic: SRR1291412-3; *E. coli* iron-starvation: SRR1168133-6). Read quality was visualized by FastQC [174] and cutadapt (v1.3) trimmed any low-quality read ends [173]. BWA-mem (v0.7.12-r1039) mapped the reads to the appropriate reference genomes (*Z. mobilis* ZM4 ASM710v1; *E. coli* K-12 ASM584v2) [175]. SAMtools sorted the aligned reads by name (v0.1.19-44428cd) [176] and BEDTools (v2.25.0) calculated per-nucleotide genome coverages for each strand (+/-) of each sample [183]. BEDTools also generated intergenic region coordinate lists from each reference genome. An R script produced comma-separated files describing the per-nucleotide read counts in the intergenic regions on each strand. Biological replicates (when available) were averaged to make a single file for each growth condition.

6.3.2 sRNA Prediction with sRNAscout

sRNAscout was developed to predict sRNA-containing regions from these intergenic region coverage files (**Figure 4.2B**). First, it normalizes coverage files by read count. Then it filters by minimum expression level (in at least one condition) and differential expression level (between the two conditions). sRNAscout extracts consecutive nucleotides of sufficient length meeting these criteria as sRNA candidate peaks. Any peaks less than 200 nt are expanded outward and combined with any neighboring peaks to yield sRNA candidate regions. Appropriate values for these criteria vary by dataset according to overall depth of coverage, read quality, and rRNA depletion. We suggest 15 nt length for minimum expression and differential expression. For the levels

of expression and differential expression, we suggest the averages of each across the dataset. **Table 4.1** shows the values used in this study.

6.3.3 Scoring sRNA Candidates for Phenotype Effect with sRNAphenoscore

sRNAphenoscore (**Figure 4.2C**) identifies which sRNA candidate regions most likely impact the phenotype of interest. BEDTools added the predicted regions from sRNAscout to the reference genome and HTSeq (v0.6.1p1) counted the reads mapped (BWA) to each sRNA candidate region [177]. DESeq2 calculates an sRNAscore for each candidate region from the absolute value of the log2 fold change between the conditions [178]. For each sRNA candidate region, IntaRNA (v2.2.0) predicts mRNA targets with most favorable binding energies in the -200 to +100 nt regions [110]. To save computational time, we precomputed the accessibility data for each genome. sRNAphenoscore calculated an mRNAscore for each sRNA candidate region: the sum of the top 50 mRNA target energies multiplied by their log2 fold changes by DESeq2. The sRNAscores and mRNAscores for all sRNA candidate regions were normalized by the max of each, yielding a value between 0 and 1. sRNAphenoscore sums the sRNAscore and mRNAscore for each sRNA candidate to produce the final phenoscore for each.

6.3.4 sRNA Candidate Overexpression Strain Development

To generate sRNA overexpression strains, GenScript synthesized and cloned each sRNA into the pBBR1MCS2-P_{gap} plasmid [171] between NheI and BamHI. Each were transformed into *E. coli* DH5α and *Z. mobilis* 8b by electroporation. *Z. mobilis* 8b grew in RMG media (glucose, 20.0 g/L; yeast extract, 10.0 g/L; KH_2PO_4, 2.0 g/L; pH 6.0) at 33°C. *E. coli* DH5α (used for plasmid construction and manipulation) grew in LB media at 37°C. Strains containing the plasmids were cultured with 50 μg/mL kanamycin for *E. coli* and

with 350 μg/mL for *Z. mobilis*. Transformants were screened by colony PCR and sequence-verified by Sanger sequencing.

6.3.5 Evaluating Strain Performance

Biological triplicates of each strain were grown in 5 mL RMG seed cultures with appropriate antibiotics at 33°C for 48 h. Cells were distributed in technical triplicates into Bioscreen C (Growth Curves US, NJ) plates with RMG (with and without 8% ethanol or 12 g/L sodium acetate) such that each well had a total volume of 300 μL and initial OD_{600} of 0.05. The Bioscreen C measured the turbidity with the wideband filter (420-580 nm) every 15 min for 24 h. The cultures grew at 33°C with 5 s of low-speed shaking before each measurement. EZ Experiment (Norden Logic Oy, Helsinki, Finland) operated the Bioscreen C a custom MATLAB script calculated exponential growth rates [172].

Appendices

APPENDIX A: SUPPLEMENTARY TABLES

Table A.1: List of 101 initial 5'UTR candidates and their features.

ID	Gene ID	Product	Start	End	Length	Counts (approx)	Rfam	Predicted Sequence	RT-PCR Conf	RACE Conf
1	zmo0172	Thiamin biosynthesis protein	155229	155392	163		TPP	agaggcatccccgggggggccgtataa atacggctgagaatgagctgattgctct aacccgtcgaacctgatccggcttaac accggcgtagggagggaaggcatca gaccttgcttcttatccgtcggccgaaa gccttttccggtcggcctgatggttga cc	Y	Y
2	ZMO0979	TonB-dependent receptor	994336	994480	144	12	Cobalamin	CGGTTTGGAGGCTGTTT TGCTGTGAGTAACAGA GGGGGGATAGAT	Y	Y
3	zmo1000	5-methyltetrahydropteroyltriglutamate	1014898	1015107	210		cobalamin	ttgcctctcgatcggagagcaatgagg aaggattaaggttctttgtcattggcaaa gctaagagggaaactggtgcgaaaga attttcaaagccagtgctgcccccgca actgtaaacggcgagcaaagatcaaa atgccactgatattattatcgggaaggc tgatcggacgcggtgacccgtcaagtc aggagacctgccttaaacca	Y	Y
4	ZMO0547	chloride channel protein	542151	542047	136	60	crcB	TAATATAGACAAGGTA ATGGAATCTACCTGATC CTTTCTATGGATCGAAC CGCCCTTTAGGCTGATG ACTCCTGCTCAAACATA TTTTTATGATTGAGGAG TCAT	Y	Y
6	ZMO0056	glucosamine/fructose-6-phosphate aminotransferase	52857	52794	57	40	no hit	ATTGCCCCCATCTCAAA GACAGTCCGCCTTTCAA AATAGATATCACAAAA TCGGGAAACAGAAT	Y	Y
7	ZMO0376	ATP-dependent protease La	383141	383033	33	13	no hit	AAATTCTTTTCCCAAAT AAGTTATTTTCCTGATA GACCCCCGTGATTGAAC CCAGAAAGATCGACAC CATATCTATGAAAAGGT GTGAGCATTTTGCCGCC CCAGTGGA	Y	Y
8	zmo0546	sulphate transporter	538383	538564	182	53	no hit	ttggtcacattgtcggcaatattcctgat attttgacgatcgggctcagccttgcta aatcccaaaaccttgatcttgcccaaaa aactgaattgatatttcccatcgcagatt ttccttcgaggaagccggataaataga gctaaaatctttcggtgaaattttgagct ctgataaaagat	Y	Y
9	ZMO0660	dnaK molecular chaperone DnaK	651270	651311	41	1100	no hit	CATTCCTGGGGAAGGA AGGTAGGACATGGGTA AAGTTATAGG	Y	Y
11	ZMO1069	molecular chaperone DnaJ	1079439	1079388	51	70	no hit	TGGCCAATTTTTCGGTC GGGGGTATTCTCATACA GGCAGAACAAGAAATA TT	Y	Y
12	ZMO1139	acetolactate synthase large subunit	1155694	1155743	49	17	no hit	GAAAAGGAAATAAAGC ATGTGTCGCCAACTCAG CGGCGCCGCCATTGTTC	Y	Y
13	ZMO1142	thioredoxin reductase	1160839	1160749	90	50	no hit	ATAATGCCGTTTTTATT CTGCTGTAGAGTGACAA TCTTTTTATCTCATTTCC CTTTTTGAAGGAGCGCT GGTGATGAGTGCTGATC CTATA	Y	Y
20	ZMO0709	phosphoribosylaminoimidazole synthetase	705230	705175	55	50	no hit	AGAAATCCAAGTCTGAT GGTGTAGGGCTTTTTTT ATGACTAATAACAGCA CGCAAA	Y	Y

25	ZMO1137	phosphoserine phosphatase SerB	1154675	1154609	66	150	no hit	CTGTTGGGCCTGCGATG AGGAGAACTGGATGTTT TTTAGGAGATTCGGATA TGCTCATTGCTACACT	Y	Y
28	ZMO0187	3-deoxy-7-phosphoheptulonate synthase	173424	173457	33	34	no hit	TTGAAAGGAATAATAA TGGTAACAGACTGGAC GC	Y	Y
30	ZMO0369	glucokinase	373612	373646	34	70	no hit	TTTACCTGTTGGGTAGC CTTCTGATTTTAGAAAG G	Y	Y
31	ZMO0405	ATP-dependent Clp protease ATP-binding subunit ClpA	405284	405322	38	200	no hit	CGATGCATCAGGAATA GGGAGTGGGAATATGC CATCTTT	Y	Y
36	ZMO0937	aromatic amino acid aminotransferase	955038	954999	39	20	no hit	TTTATCGAGGAATAGAA AGAGGGATATTATGAC CGCCTAT	Y	Y
39	ZMO1179	(uracil-5)-methyltransferase	1198318	1198272	46	28	no hit	GCTTGGCATCAAAGGG AGACGGTGTGACTGAA GATGGGCAATTTGTT	Y	Y
40	ZMO1275	threonine dehydratase	1290139	1290188	49	45	no hit	AGATTAAAAGTAATAA CGCCGGAAAAACGTTA TTGCCCAGACGCAGGAT	Y	Y
46	ZMO0689	oxidoreductase domain-containing protein	685110	685057	53	30	no hit	CCGCACAGGTGAGAAC CACGACGGATCTTCTCT GAATTGTTGGTTAGTTA AGAA	Y	Y
52	ZMO0131	metallophosphoesterase	120303	120271	32	6	no hit	ACGAGGAAATAGGGTA GAATCCATGATCTCTTT	Y	Y
53	ZMO0748	cysteine synthase	746019	745977	42	500	no hit	AGACAATCATTTTTAAG GACAGGAAAAATGGCA AAAATCAATA	Y	Y
57	ZMO1034	calcium-binding EF-hand-containing protein	1048489	1048367	122	31	no hit	CTTATTATAGGATAATC AGACAATATTCATAAAC TGCCTGATACCCATCAT CTTAGAGAAAATATTGG GATCATATCACTGTAAA ACAGTTTTCCACTTCTC TCTTGGAGTAAGGTCAA AATA	Y	Y
58	ZMO1113	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	1126999	1127156	157	970	no hit	AAAAGGTAATTGAGGC GGCTCGCAAACCACCA CGCCAAAGGGGGCTGG GTCGCTTGGAAGAGAA TAGAGGTTTCAAATGTC GAAGAATGGTAGACCT GACCGCATTGTTGTCGT GGGTGGCGGTGCCGGT GGATTGCAGTTGGTGCG TGCTTTGGGGG	Y	Y
65	ZMO0347	RNA-binding protein Hfq	340877	340965	88	150	no hit	CATCTCTTGAGAGACTC TTCACATAGGCGTGCAT CCATATTTGCTGGACGT GGATGTTTTGTAATGAT TTTGTCCAGCCAACAAA CCGG	Y	Y
68	ZMO1198	5-aminolevulinate synthase	1223191	1223264	73	15	no hit	ATCTATTCTTCCTTCTTA TAGTCAGGCTGTCACAA AAGCCTCTTTCGTCAAA AGGAGTGTCATCCTTTG GATTA	Y	Y
69	ZMO1478	6-phosphogluconolactonase	1499833	1499700	133	14	no hit	AAGGCAGAAATGCCGA TAGTCCTTAATGTCTGG CAGGTATCAATGGAGA AGAAGATGACCGAAGC CGAATGGTGGGAATTTG AAAATGTCGAAGCTAT GGCAAAACAGATTGCC GACGACATCGAATTTAT TA	Y	Y
70	ZMO0275	ABC transporter	280264	280186	78	11	no hit	TTCTGCCTTTTCCCCGA TATGGATTTGAGAATGA CTTCTGCCCGGAAAACC AAGGACAGCCTTTCTAC GGAAGCCTATG	Y	Y
79	ZMO1399	fatty acid hydroxylase	1411204	1411288	84	160	no hit	TTCGGTTCCGATTGTGA CCCGATTTGACGGGTGT CTTTCGGATCTGCGAGG	Y	Y

								ATTATGACGATGAACAC AACTGATGCCAAGACA		
83	ZMO0367	glucose-6-phosphate 1-dehydrogenase	370211	370260	49	200	no hit	CTAAAAGGATTCGGCCT CTGTTTTAAGGACGAGA ATAATGACAAATACCG	Y	Y
89	ZMO1412	MucR family transcriptional regulator	1425276	1425220	56	32	no hit	AGATAGAATATTCTCCC ACTTTTTTCCTAAAAAA GGATTGTTCTATGAGCG ATGAC	Y	Y
91	ZMO1048	phosphate ABC transporter inner membrane subunit PstC	1062465	1062535	70	180	no hit	GGAAAAAGATGATAGC TGTCAGAAGAGAGTCG GATAAATTTTATGAGTG ATTCTGCCTTGTCCCAG CCTTC	Y	Y
92	ZMO0140	protein tyrosine phosphatase	128038	127939	99	4	no hit	CATAACGATGGATGAA TCCGTTTAATCCTTTTAT CCGATCCGAAACAAAG CCTTTTTTTGCAGGAGT CCTAAACCCGATGACCT CGACCCCGACCGTGCT	Y	Y
95	ZMO0366	sugar transporter	368629	368728	99	30	no hit	CCGATACTCGGCGATTG TAAGATTTACAGATTAA GGCGGGAGAGGAATCG CCATGAGTTCTGAAAGT AGTCAGGGTCTAGTCAC GCGACTAGCCCTAATC	Y	Y
96	ZMO1432	fusaric acid resistance protein	1448624	1448647	33	130	no hit	TAAACAGCATCGCAGTC GCCTAT	Y	Y
99	ZMO1612	toluene tolerance family protein	1652781	1652835	54	90	no hit	TTTAGCCTTGAAAACCG ACTTCTCTGACACCGTC CGGGAAAGGAGAATTA TGAA	Y	Y
10	ZMO0732	DNA-directed RNA polymerase subunit beta'	729699	729732	33	200	no hit	CTTTCTGAGGGATGAAG AATGAACGAATTAGCT A	Y	-
15	ZMO1324	HPr kinase	1341681	1341765	84	30	no hit	ATCGCTTTTGACTGTGA TTATCTTTGTAATGTCA TGGGCTAAAGCGTTCTA ATTGAGATGTCTTCTGT AACTATTGAAAATACC	Y	-
16	ZMO1480	electron transfer flavoprotein alpha/beta-subunit	1501832	1501770	62	30	no hit	CTAACTCCAGCATGCTT CATAAATCGGCAAAGG GAGATTCTGATGAAAGT AATAGTGCCCGT	Y	-
17	ZMO0077	(dimethylallyl)adenosine tRNA methylthiotransferase	68494	68464	30	60	no hit	TTTTTTATGAACCGACC TTCTTTAAAGCATT	Y	-
18	ZMO1360	thiamine pyrophosphate protein TPP binding domain-containing protein	1375102	1375030	72	187	no hit	CATAGTGTTTTGAATAT ATGGAGTAAGCAATGA GTTATACTGTCGGTACC TATTTAGCGGAGCGGCT TGTCC	Y	-
41	ZMO0247	hslU ATP-dependent protease ATP-binding subunit HslU	239553	239594	41	270	no hit	CTTGCTGGAAAAAGATT GTATGACTGATGCCCTT ACGCCCAA	Y	-
45	ZMO1682	aspartate aminotransferase	1726052	1726109	57	14	no hit	TTCAATTAGATAAAGAA GAAATCAAATGACGAC TGATTACTCGAAATATA AAAATCT	Y	-
54	ZMO0918	catalase	934776	934849	73	11	no hit	AAATCCCAATCAGGTA AGAGGGACGTTATGAC TAGACCCAATCTTCAGA CTGAATCAGGCCGCATT GTTGGTGA	Y	-
55	ZMO1029	ABC transporter-like protein	1043210	1043302	92	10	no hit	CTGAATAGCCAATAACC CTGACGTCTTTTCTGTTT TCCGGTTCAAAGGGAG AAAATTTTGTCCGCAGC TATTCAGCTTGCCCATC TTTCCAAA	Y	-

74	ZMO1770	argininosuccinate lyase	1813959	1813898	61	80	no hit	GGTAATGACCCAGTCG GGAGACAGATGATTAC TGAGCGTCAGACAGATT CCAATGCGATGT	Y	-
77	ZMO1033	cyclopropane-fatty-acyl-phospholipid synthase	1046361	1046415	54	24	no hit	GTTAAAGCATGTAAACT CTGACCATGTGGCTACT TGATAGATTCTTAAAAA AGCT	Y	-
78	ZMO1008	FAD linked oxidase domain-containing protein	1025610	1025677	67	9	no hit	TCTTAAAAGGGTATTTC ATGGAGATTTGCTATGT CCCGTCAGCCAGAAAA GGCTTTGCTTGATGCAT T	Y	-
84	ZMO1556	glutamate/cysteine ligase	1586528	1586568	40	86	no hit	GACTATATCTTGTGTTA TGAGCACGCGTCAGAC ATCCAGC	Y	-
97	ZMO1097	thioredoxin	1109452	1109406	46	60	no hit	ATATAAATTCATATCCC CGTCAGGAGCGATTATA ATGAGCGTTATCA	Y	-
5	zmo0861	DNA polIII subunit	868914	869065	152		Bacterial small SRP	tcataggcgatagtggaagtcgggcg gacggtgctgtcgccaacccggtcag gtccggaaggaagcagccgtaacga atttgtatcgggtcgttccggcttccacg tcttaaagatttctaaaaattccttttttctt tatccttgataatc	-	-
14	ZMO1215	hypothetical protein	1239953	1240024	71	60	no hit	GATTACCTCGGTAAGAC TGAACCGTATCCGACCC GATCATAGAATAACTA AGATGAATGGCAGCAT GACAGC	-	-
19	ZMO0017	Fmu (Sun) domain protein	18960	19010	50	40	no hit	CATAACCTTCGGTTGTC CATTCGGGCGGGAATTT TATGAATACTACACGG	-	-
21	ZMO1184	electron-transferring-flavoprotein dehydrogenase	1205615	1205564	51	150	no hit	AAACCATAAGGATTAT GGGGGAAATGTCAGAA CGCGAATCCATGGAAT ATGA	-	-
22	ZMO1708	pyridoxine 5'-phosphate synthase	1756860	1756904	44	30	no hit	AAACAAGGCTCAAGAG GACGAAATGACCAAAA GAATAAGGCTTG	-	-
23	ZMO0027	phosphoribosylaminoimidazolecarboxamide formyltran	29250	29292	42	10	no hit	TGCCGCCCAGATTAGTC GGAAAAACTGACAGGA TAAGGCATC	-	-
24	ZMO0914	methenyltetrahydrofolate cyclohydrolase	930011	929969	42	14	no hit	TTCTTTTAATATTTTTGT TGTAAGGATTGTTCATG ACTGAGGC	-	-
26	ZMO1485	deoxyguanosinetriphosphate triphosphohydrolase	1510314	1510385	71	250	no hit	ATCTCTTTGGAATAGAC GAAATGAAATGGGAAA GACTACTGAGCCACAA GCGGCTATCGGCACCTG ATTAT	-	-
27	ZMO0120	dihydroorotate dehydrogenase	107617	107652	35	3	no hit	CTCGGGAATCATTTGGG AGGCCAGAGCATGAAA ACC	-	-
29	ZMO0215	5-formyltetrahydrofolate cyclo-ligase	208225	208274	49	54	no hit	ATAATGAATATTTTTCT GTGCTTGGATGGAGAA GCCTGTGTCCGTTGCGT	-	-
32	ZMO0443	nrdF ribonucleotide-diphosphate reductase subunit beta	438396	438341	55	200	no hit	TCATTTCAGACATCCTT CCTTAATATC	-	-
33	ZMO0820	phosphoribosylformylglycinamidine synthase II	822075	822036	39	30	no hit	ATTAAAGATAGGTTGG GTGCTGACAGAAAGGT TACACTTC	-	-
34	ZMO0903	2-isopropylmalate synthase	913395	913287	108	65	no hit	AACGTTGTTGTTCTGGT GATCCAATCACCGGAC AAGGCAGAACGAAGTT TTAGGGATAATCATGTT ACATAATCCGGCTACAA	-	-

								AATATCGTCCTTTCCCT GCCATCGAT		
35	ZMO0923	argJ bifunctional ornithine acetyltransferase/ N-acetylglutamate synthase	942907	942871	36	28	no hit	GTCTATTGTGATTGAAG GATCTTTTTGTATGATC GCC	-	-
37	ZMO1020	Orn/DAP/Arg decarboxylase 2	1037070	1037024	46	16	no hit	TCTTGCCACCCTACCCT CTGGAGGTCCCTTGAAT TGCACAATCATCA	-	-
38	ZMO1052	phosphoribosylaminoimidazole-succinocarboxamide synthase	1065453	1065513	60	45	no hit	TTATTAAGCGCCTTGCC TTTTTCGGCACAGGATC TGTCCTGTTCCTTTTTCG GAGAAGCTA	-	-
42	ZMO0309	short-chain dehydrogenase/reductase SDR	310125	310170	45	34	no hit	GGCTTTCGGCCTAATAG GACTTTTACGACCTGTA TTTTCTGAAAAT	-	-
43	ZMO0512	dihydrolipoamide dehydrogenase	510820	510854	34	24	no hit	TTACAAAGGAAAGAGA ACTTAACTCATGGCTGA TC	-	-
44	ZMO1181	N-formylglutamate amidohydrolase	1200963	1200868	95	16	no hit	AATAAAAACCTTATAG ATGATATAAGATCGTCT ATAAGGTTTTTTGTGTT TTTCTTAAGCGTAAATT GATAGAGGTTGAGTCGT GTCTTATCAAGT	-	-
47	ZMO1236	alcohol dehydrogenase	1260469	1260362	107	200	no hit	TCCCTATATTCGCAAGA TGTATGTCAGCCGCAGA TTTGTCGACTGACCTCT ATCTCTCCGAGATATAT CAACAAAAGGTAGTCA CCATGAAAGCAGCCGT CATAACT	-	-
48	ZMO0976	aldo/keto reductase	993258	993201	57	29	no hit	CTTTTTTAAAAAATGTG GAGAGTTTCTGATGAAC ACTTCTACGCAAAAACC CGCTCAT	-	-
49	ZMO1522	TonB-dependent receptor	1545002	1545044	42	19	no hit	TTGGGGACGGTGCATAT CTCAAAAAGGGGGAAC AATGATCAG	-	-
50	ZMO0130	Acid phosphatase	119200	119168	32	55	no hit	TAACAAGGAAAGACTG ACGACATGATAAAAGT C	-	-
51	ZMO0031	TonB-dependent receptor	36817	36773	44		no hit	TAGAGGGCAGCGGCAT TTGGCCTTTGGAAGGCT GTCGGGCGAAA	-	-
56	ZMO1410	bacterioferritin	1423489	1423534	45	300	no hit	TATCCTAATTAGGGTTG TAAAAAGCAAAAGGTA GCATCCATGAAG	-	-
59	ZMO0753	glutaredoxin 3	749452	749512	60	200	no hit	CCAGTAACATTTAGTCG GTTTTCTCCTATCTTAG GATTGGAGATGACGCG GCCTTCGCTTC	-	-
60	ZMO1639	integral membrane protein MviN	1680356	1680448	92	30	no hit	GCCGCTTTTGAGCGGCT TTCTTCACTTTTATCCA GATGACGATATGACAA ATAAAACTCAGCCTGCC TCCGGACGATTTCATCT TTTCAAAG	-	-
61	ZMO1350	molybdopterin binding domain-containing protein	1363537	1363602	65	40	no hit	TTTTTGGATTCGGAATA TCTGCCGCATCAGAGGA AAGAGGAAAAGGACGG ACTGAAAATGACAAC	-	-
62	ZMO1576	short-chain dehydrogenase/reductase SDR	1607650	1607735	85	110	no hit	ATTTCTATCATCCGGTG CGAAGAATTTTGTCTTG GATTTTTTGAGCATAGA GGCTGATTATGAACCAG AATATCCGCAATAAAAT	-	-
63	ZMO1705	thioredoxin	1756074	1756028	46	2000	no hit	ATATCAAAGAAACGGG CAATAAAAAGGACGCT GATTGTGTCAACCT	-	-
64	ZMO1385	toxic anion resistance family protein	1396885	1396933	48	61	no hit	GATTTTACAGATGGATA GTTGGGGAGTGCCGCTT GATGGTTACAGAAA	-	-

66	ZMO1756	gluconate transporter	1799573	1799699	126	60	no hit	TGCAAAATAGACTGAA TGCTGGCTTTCCCATAG ATGGAATTTCTGCATCT TGCGTCAGGTCTTCATA AGGCCTGACCAAAAAA AGAACGGATGAGTAAC CATGCACTCTAATGGAA CAATGCTCCT	-	-
67	ZMO1222	3-oxoacyl-ACP reductase	1246907	1246767	140	42	no hit	CAGCCCCTCTTCTGGTT CTATACTCTAACCAGAC AGGGGGCAGGAGAAAA TCAATGTTTGATCTTAA TGGCCTGACTGCTGTTG TAACGGGTGCCTCTGGC GGCATTGGCTCTGCCAT CGCCAAAGCCTTGGCCG ATCAA	-	-
71	ZMO0778	acriflavin resistance protein	770907	770992	85	38	no hit	GCAAATCAAACAGACA TCCGTATCAACGATCTG GAACACCGCTTATGCTC GCGCTGGTACGCATTGC TCTCCGTCGCCCTTATA C	-	-
72	ZMO1673	aldo/keto reductase	1718309	1718190	119	51	no hit	GCAAGAAAGGATTGGA GTCCTTTATGGTGCCGG TTAAAAATGTAGTGCTG AATGACGGGCATAGTA TGCCAGCCATCGGTCAA GGCGTTTACCAACTCAA TCCCGCACAAGCTGAA AGC	-	-
73	ZMO1497	amino acid permease	1524545	1524497	48	16	no hit	CGGGGGATAATGAATG ATATTCGGGCGTGTGAA GCCTCTCGACGCCAT	-	-
75	ZMO1107	AsnC family transcriptional regulator	1122551	1122592	41	75	no hit	GCCGGAAGATTAAATA GAAGGACTAAGTCCGT GACGGCTAAT	-	-
76	ZMO1407	aspartate-semialdehyde dehydrogenase	1419773	1419653	80	25	no hit	TAATACGGTTTAGACTG ATAAGTATAAGTGGAG AAAATAATGGGATATC GGGTCGTAGTCGCAGG CGCTACCGGTAACGT	-	-
80	ZMO0179	fructose-bisphosphate aldolase	164659	164697	38	23	no hit	GGGAAGGAGCGCTGTG ATGTCTATTACTCCCGA AGTAAA	-	-
81	ZMO1235	Fur family ferric uptake regulator	1259164	1259117	47	39	no hit	AAGGATATTAGCGTTTA TATATAATAGAAGGAA ATCTGGCCTTGGGT	-	-
82	ZMO1411	Fur family ferric uptake regulator	1424587	1424553	34	20	no hit	TTGACGGAGAAATGTTG TGCGGCAGATCAATATT	-	-
85	ZMO1599	hopanoid biosynthesis associated RND transporter like protein HpnN	1638166	1638280	114	200	no hit	GATCCAAGTTATCTTTT ATTCTGAATCGTCCGGT TGTAACAGCTAAGGCG CTTGGGCTAGCCATACG GATTGATCTCGTGATAT CGGACACATCATGCATC GTTTTCTTGTAAC	-	-
86	ZMO1596	iron-containing alcohol dehydrogenase	1634585	1634653	68	100	no hit	AATGGTTGTTTTCGGGT TGTTGCTTTAAACTAGT ATGTAGGGTGAGGTTAT AGCTATGGCTTCTTCAA	-	-
87	ZMO1771	iron-containing alcohol dehydrogenase	1815437	1815373	64	130	no hit	CGCATAAGAAGGCGGA AAAGACTGATGCTCAAT TTTGATTATTATAATCC GACCCATATTGTTT	-	-
88	ZMO1233	mannose-1-phosphate guanylyltransferase	1256501	1256380	121	190	no hit	AGGGAGCCTCGGCTCCC GACAATAGTAAAGAGA TCATATATAATGCTACA TCCGGTTGTTTTGTGTG GTGGTTCAGGTACACGT CTTTTCCCGTTATCCCG CCGGAGCCATCCCAAA CAAC	-	-
90	ZMO0178	pgk phosphoglycerate kinase	162649	162720	71	110	no hit	GTAAGGAGGCTGTCTCC GTTATCAGATCCGCTTT ATGATATAAGGCGGCC AAAAGGAGGATATAAT GGCTTT	-	-
93	ZMO1606	pyruvate dehydrogenase (acetyl-	1648817	1648755	62	14	no hit	GATTTATCCTGTCCGTC AGAAAATTTTTAGTCTT TTTTGACAGGAGGCCTC TGCATGGCCAA	-	-

		transferring) E1 component subunit alpha								
94	ZMO0322	riboflavin biosynthesis protein RibF	318711	318742	31	40	no hit	CTTATGTAAATCGAGAC TATGACCTGTCTTTT	-	-
98	ZMO0765	thrS threonyl-tRNA synthetase	756950	756992	42	450	no hit	TCTTTGAAATAAAAGAA AGACGAGAATGGCTGA GAAGATCAG	-	-
100	ZMO0054	transcriptional regulator, MarR family	49298	49373	75	2	no hit	TTTGAAGACGGCTTGCG TGAGACGCGTGACAAG CCCAAAAAGGGAATCC TATGAGAGATCACGTA GATTTTGTGG	-	-
101	ZMO1289	Transglycosylase-associated protein	1303429	1303320	109	110	no hit	AATTCTTTTTTTTAAGG GAGTTTGGTAATGGGGC TCATTTCTTGGCTTATC GTCGGCGGTATTATTGG ATGGCTCGCTGGACTTA TTATGAAAGGCAGCGG CAGCGGCG	-	-

Table A.2: List of final 5'UTR candidates and their features.

Gene ID	5'UTR Sequence Confirmed by 5' RACE	Strand	Start	End	Known riboswitches in other organisms	Gene regulation in literature	Number
ZMO0172	tccccgggggccgtataaatacggctgagaatgagctgat tgctctaacccgtcgaacctgatccggcttaacaccggcgta gggagggaaggcatcagaccttgcttcttatccgtcggccg aaagccttttccggtcggcctgatggttgacctgcaagaag gatgcataa	+	155236	155410	TPP riboswitch in E.coli		RSE#1
ZMO0979	ggaaatttttttgcataggggtttccttcgagtgaaggaaaattg ggaacaaggtgcaaaaccttggctgccccctgcaactgtaaa cagttgaaacgccaaaaagccactgaatctattcgggaagg cggttgtttcgatgctgtgagccaggagaccgaccctatgta atcgttccacgacaggacagggcgttcccgtcgaacgagg taaattctcctcgcacggacgataatagtgaccttcaagcca atcggtttggaggctgttttgctgtgagtaacagagggggga tag	+	994310	994604	AdoCbl	Downregulated in Acetate and Xylose stress	RSE#2
ZMO1000	aaggattaaggttctttgtcattggcaaagctaagagggaaa ctggtgcgaaagaattttcaaagccagtgctgccccccgcaa ctgtaaacggcgagcaaagatcaaaatgccactgatattatt atcgggaaggctgatcggacgcggtgacccgtcaagtcag gagacctgccttaaaccaagtcatccactatcggttttgaaaa ctggaaacaggcttttaaaactggtttcttttcggggttgaag agaaaa	+	1014925	1015181	metE in E.coli/Bacillus	Upregulated in Xylose	RSE#3
ZMO0547	gaatgaccatttccattatccactgacttgacttttggcattattt tggccttaatatagacaaggtaatggaatctacctgatccttttc t	-	542113	542203	Fluoride riboswitches		RSE#4
ZMO0056	atctcaaagacagtccgcctttcaaaatagatatcacaaaatc gggaaacagaatt	-	52793	52848	glmS ribozyme		RSE#5
ZMO0376	aaaaggtgtgagcattttgccgccccagtggagttgtg	-	383027	383064	-		RSE#6
ZMO0546	aattttggtcacattgtcggcaatattcctgatatttttgacgatc gggctcagccttgctaaatcccaaaaccttgatcttgcccaa aaaactgaattgatatttcccatcgcagattttccttcgaggaa gccggataaaatagagctaaaatctttcggtgaaatttgagct ctgataaaagatggcaatctcaattcctagcatttttaaggata cccg	+	538379	538600	ABC transporter family	Upregulated in 5% Ethanol and Xylose stress	RSE#7
ZMO0660	gtagacttgttccaagggcaagctttattttagaagagtaaa ttttgtcattcctggggaaggaaggtaggac	+	651220	651294	-		RSE#8
ZMO1069	tttgctggtggacagaacccggcttatagaccttctggcgtat cacctctgttgatttttctttgtcatttgcctctttttcctaaagaaa ctatattcttggttttccgactttggccaatttttcggtcgggg gtattctcatacaggcagaacaagaaatatt	-	1079388	1079553	-		RSE#10
ZMO1139	cccttatcaaggctacaaactgaaccatgacaatcatcgcga ttggatcatggggtttcgaaaaggaaataaagc	+	1155635	1155709	-	Upregulated in Xylose	RSE#11
ZMO1142	tctgctgtagagtgacaatcttttatctcatttccctttttgaag gagcgctggtg	-	1160767	1160823	-	Upregulated in Xylose	RSE#12

ZMO0709	acatgcaacctttgaacagttcagcttagaaatccaagtctg atggtgtagggctttttt	-	705197	705258	-	-	RSE#17
ZMO1137	ggagaactggatgtttttaggagattcggat	-	1154626	1154657	-	serC regulated by glycine riboswitch	RSE#18
ZMO0187	ttttttgccagagctttctctgacagaaagaacgctcctgcc gaatagaaaactgatgaaccgcctttgcgcagcttgaaagg aataata	+	173348	173438	-	-	RSE#19
ZMO0369	Tttacctgttgggtagccttctgattttagaaaggaattatt	+	373612	373653	-	-	RSE#20
ZMO0405	agacgagcttatttaggcatntgattttatccg	+	405253	405285	-	Upregulated in Xylose/ down regulated in Ethanol stress	RSE#21
ZMO0937	aatatccctctttctattcctcgataaaaagaggttaaacgcaa tttaatccctgaaacaaggataaagcataaggcttttctgtatc tataaaatgatagcacaggataataaatttacttttattatagag ataatgtctttttctttatatccatttttatagctatattgatatataa ttatttttgtgatctgtatcatataacggataataatataattttgt tatcctctttaataaaaagaggcttccctatatcgaaatgtcac atttcgatgatattattggtttttcctagaatttttaaggg	+	955011	955326	-	Acetate/xylos e down	RSE#22
ZMO1179	gtaacagtcgcctcatgaaaaaaattatacgcttggcatcaa agggagacg	-	1198298	1198348	-	5% Ethanol up	RSE#23
ZMO1275	aatactatctgttttagattaaaagtaataacgccggaaaaac gttattgcccagacgcaggataaagc	+	1290124	1290192	-		RSE#24
ZMO0689	ctggaaaagccggcgaatcagataacagttccgcacaggt gagaaccacgacggatcttctctgaattgttggttagttaaga aagaaacaaggatt	-	685044	685140	-	xylose down	RSE#27
ZMO0131	ctggcctcatttctctcgaaatgaaaatttcaaaaaccgatatt tttgtaggttcacgaggaaatagggtagaatcc	-	120282	120358	-	Downregulat ed in Acetate and Xylose stress	RSE#28
ZMO0748	aagacgcgccctacatataaatgaacagattattctgcctag acaatcatttttaaggacaggaaaa	-	745993	746059	-	Upregulated in Acetate and Xylose stress	RSE#29
ZMO1034	aatcagacaatattcataaactgcctgatacccatcatcttag agaaaatattgggatcatatcactgtaaaacagttttccacttc tctcttggagtaaggtcaaaata	-	1048367	1048476	-	Upregulated in Acetate and Xylose stress	RSE#32
ZMO1113	aattgaggcggctcgcaaaccaccacgccaaaggggct gggtcgcttggaagagaatagaggtttcaa	+	1127006	1127074	-	xylose up	RSE#33
ZMO0347	atcccgatggcgttttttcttttcttttacaattctatgataattgaa taatatttcattgtccttttgtaataaaggatagaaaaatggattg gcctagtaaattgtttctgtgacatttgaaattttaaactatatta tcatctcttgagagactcttcacataggcgtgcatccatatttg ctggacgtggatgttttgtaatgattttgtccagccaacaaac cggagaaccggacaatggccgaaaaggtcaacaatcttca ggatttttttccttaataccttgcgcaagacccgcacaccggtg acg	+	340738	341048	-	Downregulat ed in Ethanol stress	RSE#34
ZMO1198	attcttccttcttatagtcaggctgtcacaaaagcctctttcgtc aaaaggagtgtcatcct	+	1223195	1223256	-	xylose	RSE#35
ZMO1478	atattctcaatagtctggcaatgagtcgtcaaaatagtttgaaa ctgttctaagacgacgccgttctgtaagtgagatagataaaac ctgctcaaaaggcagaaatgccgatagtccttaatgtctggc aggtatcaatggagaagaag	-	1499779	1499926	-		RSE#36
ZMO0275	atccaaaaggttgcctgtttttcaatcggttaaacgggctgcc tctttttttctgccttttccccgatatggatttgaga	-	280235	280314	-		RSE#37
ZMO1399	aataaaagcaacaaccttctgaacttcggttccgattgtgacc cgatttgacgggtgtctttcggatctgcgaggattatgacg	+	1411180	1411263	-		RSE#41
ZMO1432	ttttacaaataggcataaaacgcttttcttaacaggtgatgtct gcccgattatgtatgatatgaccgttttacactataaacattgc acaagggtaaaggacaggccactgcgcctatcgactcttgc tatgataggcgactgcg	-	1448634	1448779	-		RSE#42
ZMO0367	ttaaacttgctttggctgaatccttttgtctttttagataagtctt aaccaattatacttttttgtttacaacgatggtataaagcgggc ggacaggctaaaaacaggctaaaaggattcggcctctgtttt aaggacgagaata	+	370103	370247	-	Upregulated in Ethanol stress	RSE#43
ZMO1412	gtaaacgttgggagatcaaccgagatctttccaaaggtaaa agctttaaagcttctctggaaagatagaatattctcccactttt tcctaaaaaaggattgttct	-	1425232	1425337	-	EtOH proteomics	RSE#45

ZMO1048	gtatattttggctgtttgttgtaatattctacatcagcctaaaca gatatgaggaaaaagatgatagctgtcagaagagagtcgg ataaatttt	+	1062413	1062506	-	RSE#46
ZMO0140	agcgctacagaaaaatataggaatttcaccgccgtttttttcccaa aaaacagtgatgtataccctcgctattcagggataaaaaat ctaaaattaaacataacgatggatgaatccgtttaatccttttat ccgatccgaaacaaagccttttttgcaggagtcctaaacccc g	-	127962	128134	-	RSE#47
ZMO0366	agtttatcgccaaggattcggtttgtgatgttatttttggtcaga aactaaaataagaccaatgtttaacattgccgatactcggcg attgtaagatttacagattaaggcgggagaggaatcgcc	+	368555	368680	-	RSE#48
ZMO1612	attgattcgcccgaataaccattagagaagaatctgcatatca acaggccattntccttgaaggtgtttttaacattagccgaatc agcaaaaagccagaaaatctggattagccttgaaaaccga cttctctgacaccgtccgggaaaggagaatt	+	1652672	1652829	-	RSE#50
ZMO0732	*connected with adjacent gene rpoB*					*RSE#9*
ZMO1324	*overlap with adjacent gene*					*RSE#13*
ZMO1480	*no 5'UTR*					*RSE#14*
ZMO0077	*no 5'UTR*					*RSE#15*
ZMO1360	*only22mer*					*RSE#16*
ZMO0247	*too short intergenic*					*RSE#25*
ZMO1682	*overlap with adjacent gene*					*RSE#26*
ZMO0918	*only 16mer*					*RSE#30*
ZMO1029	*adjacent gene transcribe together*					*RSE#31*
ZMO1770	*overlap with adjacent gene*					*RSE#38*
ZMO1033	*No 5'UTR*					*RSE#39*
ZMO1008	*overlap with adjacent gene*					*RSE#40*
ZMO1556	*no 5'UTR*					*RSE#44*
ZMO1097	*only 21mer*					*RSE#49*

Table A.3: List of primers for 5'UTR study.

Primer name	Sequence	Function	Target gene
zmo0172_5U_F1	gtagggagggaaggcatcag	RT-PCR	ZMO0172
zmo0172_mRNAF2	cgaacgccatttttaccttc	RT-PCR	
zmo0172_mRNARv	cacgcaaaggatgatggtt	RT-PCR	
zmo0979_5U_F1	attgggaacaaggtgcaaaa	RT-PCR	ZMO0979
zmo0979_mRNAF2	cgactttggctttgacaaca	RT-PCR	
zmo0979_mRNARv	actgttactccgggcgttct	RT-PCR	
zmo1000_5U_F1	gatcggagagcaatgaggaa	RT-PCR	ZMO1000
zmo1000_mRNAF2	tttgggtttcccaagaattg	RT-PCR	
zmo1000_mRNARv	tgttttcagcagaggtcgtg	RT-PCR	
zmo0547_5U_F1	agacaaggtaatggaatctacctga	RT-PCR	ZMO0547
zmo0547_mRNAF2	ccttgattaaaggatgccaga	RT-PCR	

zmo0547_mRNARv	gccaccaaaaagatgggtta	RT-PCR	
zmo0861_5U_F1	ccggcttccacgtcttaaa	RT-PCR	ZMO0861
zmo0861_mRNAF2	atcgcatcctgaccaatca	RT-PCR	
zmo0861_mRNARv	atggttgattcttcctctctgg	RT-PCR	
zmo0056_5U_F1	aaagacagtccgcctttcaa	RT-PCR	ZMO0056
zmo0056_mRNAF2	caaggattaagacggctgga	RT-PCR	
zmo0056_mRNARv	tcagaaccttggaaacagca	RT-PCR	
zmo0376_5U_F1	cgccccagtggagttgtg	RT-PCR	ZMO0376
zmo0376_mRNAF2	gctttttgtcggtcgtgaa	RT-PCR	
zmo0376_mRNARv	actttgacagccaaattaacagc	RT-PCR	
zmo0546_5U_F1	cagccttgctaaatcccaaa	RT-PCR	ZMO0546
zmo0546_mRNAF2	cagcgcaaagagtccaaaag	RT-PCR	
zmo0546_mRNARv	tagataaccgcagcagacca	RT-PCR	
zmo0660_5U_F1	attcctggggaaggaaggta	RT-PCR	ZMO0660
zmo0660_mRNAF2	acagctgcgttgctgttatg	RT-PCR	
zmo0660_mRNARv	tcttaccggcatctttggtt	RT-PCR	
zmo0732_5U_F1	tttgttgacccttctgagg	RT-PCR	ZMO0732
zmo0732_mRNAF2	tgccagtacagagcatttcg	RT-PCR	
zmo0732_mRNARv	tcggtcagaagctggaattt	RT-PCR	
zmo1069_5U_F1	aatttttcggtcgggggtat	RT-PCR	ZMO1069
zmo1069_mRNAFw	gttggcgcggttataatga	RT-PCR	
zmo1069_mRNARv	gattaagatcgggatgccatt	RT-PCR	
zmo1139_5U_F1	cgattggatcatggggttt	RT-PCR	ZMO1139
zmo1139_mRNAFw	cgccattgttcttgaaacac	RT-PCR	
zmo1139_mRNARv	agatggaaggcctcgacaat	RT-PCR	
zmo1142_5U_F1	tgccgtttttattctgctgt	RT-PCR	ZMO1142
zmo1142_mRNAFw	gctgatcctatatctacccgtgtt	RT-PCR	
zmo1142_mRNARv	ttccacggaaaaagaaacca	RT-PCR	
zmo1215_5U_F1	aagatagattacctcggtaagactgaa	RT-PCR	ZMO1215
zmo1215_mRNAFw	aatggcagcatgacagcac	RT-PCR	
zmo1215_mRNARv	tttgccaaaagggctagaag	RT-PCR	
zmo1324_5U_F1	cgcttttgactgtgattatctttg	RT-PCR	ZMO1324
zmo1324_mRNAFw	cggtggttcaaacgaatcat	RT-PCR	
zmo1324_mRNARv	tcaatttttatcggtgtcgaag	RT-PCR	
zmo1480_5U_F1	cactaactccagcatgcttcata	RT-PCR	ZMO1480

zmo1480_mRNAFw	taatagtgcccgttaaacgtgtg	RT-PCR	
zmo1480_mRNARv	ttttattatcggcaatcgtcaa	RT-PCR	
zmo0077_5U_F1	tcggcttttcaactgtttgg	RT-PCR	ZMO0077
zmo0077_mRNAFw	ccgaccttctttaaagcattttt	RT-PCR	
zmo0077_mRNARv	aatttatcgcagccttcctg	RT-PCR	
zmo1360_5U_F1	ttcagacatagtgttttgaatatatgg	RT-PCR	ZMO1360
zmo1360_mRNAFw	gagttatactgtcggtacctatttagc	RT-PCR	
zmo1360_mRNARv	atcgattttagccggagctt	RT-PCR	
zmo0017_5U_F1	accttcggttgtccattcg	RT-PCR	ZMO0017
zmo0017_mRNAFw	gtcccaaagggaaagacagc	RT-PCR	
zmo0017_mRNARv	taaacgccatgaaccagacg	RT-PCR	
zmo0709_5U_F1	tccaagtctgatggtgtaggg	RT-PCR	ZMO0709
zmo0709_mRNAFw	aacagcacgcaaaaatacagtt	RT-PCR	
zmo0709_mRNARv	gtaatcgccatcggcataga	RT-PCR	
zmo1184_5U_F1	ggatgattgttatataaaccataagga	RT-PCR	ZMO1184
zmo1184_mRNAFw	atgtcagaacgcgaatccat	RT-PCR	
zmo1184_mRNARv	ttttcgtcatacaaaatatctgctg	RT-PCR	
zmo1708_5U_F1	ccgcttaaaacaaggctcaa	RT-PCR	ZMO1708
zmo1708_mRNAFw	aagaataaggcttggggtcaa	RT-PCR	
zmo1708_mRNARv	attcgataaccggcgcatt	RT-PCR	
zmo0027_5U_F1	cgcccagattagtcggaaa	RT-PCR	ZMO0027
zmo0027_mRNAFw	ttaaaaccgcactgctttcc	RT-PCR	
zmo0027_mRNARv	gataatctttcgggtcggtca	RT-PCR	
zmo0914_5U_F1	taatatttttgttgtaaggattgttc	RT-PCR	ZMO0914
zmo0914_mRNAFw	gtaacgtcgttcctggcttg	RT-PCR	
zmo0914_mRNARv	accaaagcatttttgcctttt	RT-PCR	
zmo1137_5U_F1	gcgatgaggagaactggatg	RT-PCR	ZMO1137
zmo1137_mRNAFw	atcaagctcggtcgcttct	RT-PCR	
zmo1137_mRNARv	tgatccctggggttaaatga	RT-PCR	
zmo1485_5U_F1	ttgaataaatctctttggaatagacg	RT-PCR	ZMO1485
zmo1485_mRNAFw	tcaagatgcagaccggatta	RT-PCR	
zmo1485_mRNARv	atttcaagctgggtcagcaa	RT-PCR	
zmo0405_5U_F1	gcatcaggaatagggagtgg	RT-PCR	ZMO0405
zmo0405_mRNAFw	atgaacatgcttcgcaagtg	RT-PCR	
zmo0405_mRNARv	gctgggtcttttcttcttcg	RT-PCR	
zmo0443_5U_F1	cattttcagaatcggcagga	RT-PCR	ZMO0443

zmo0443_mRNAFw	cagcagcttcattggcttc	RT-PCR	
zmo0443_mRNARv	gctgccgacatcaaaggtt	RT-PCR	
zmo0820_5U_F1	tgggtgctgacagaaaggtt	RT-PCR	ZMO0820
zmo0820_mRNAFw	gacaacgcctgaaaccaaat	RT-PCR	
zmo0820_mRNARv	tcccccgaagaatatgacac	RT-PCR	
zmo0903_5U_F1	cgttgttgttctggtgatcc	RT-PCR	ZMO0903
zmo0903_mRNAFw	gatttacccaatcggcagtg	RT-PCR	
zmo0903_mRNARv	cacctgcgatagccagtttt	RT-PCR	
zmo0923_5U_F1	gccagtctattgtgattgaagg	RT-PCR	ZMO0923
zmo0923_mRNAFw	cgcaatgatctcagcctgt	RT-PCR	
zmo0923_mRNARv	tggcattagccatatgttcc	RT-PCR	
zmo0937_5U_F1	tttatcgaggaatagaaagagggata	RT-PCR	ZMO0937
zmo0937_mRNAFw	tcgtgaagatactcgcgaaa	RT-PCR	
zmo0937_mRNARv	ccagtattggcaatgatctgaa	RT-PCR	
zmo1020_5U_F1	gtcttgccaccctaccctct	RT-PCR	ZMO1020
zmo1020_mRNAFw	ttgcgggaaagacgatgtat	RT-PCR	
zmo1020_mRNARv	cagggtcaatgccgaattt	RT-PCR	
zmo1052_5U_F1	attaagcgccttgccttttt	RT-PCR	ZMO1052
zmo1052_mRNAFw	ccggcaccattatccagtat	RT-PCR	
zmo1052_mRNARv	tgcatttcttcctgcgtagc	RT-PCR	
zmo1179_5U_F1	cttggcatcaaagggagac	RT-PCR	ZMO1179
zmo1179_mRNAFw	cggttgccgtttacaatatg	RT-PCR	
zmo1179_mRNARv	tcaattccttgatctgctaacg	RT-PCR	
zmo1275_5U_F1	ttaaaagtaataacgccggaaaa	RT-PCR	ZMO1275
zmo1275_mRNAFw	tatcaaacaggcgcatcaac	RT-PCR	
zmo1275_mRNARv	agattgtctttttcagccaattc	RT-PCR	
zmo0247_5U_F1	cttcttgctggaaaaagattgt	RT-PCR	ZMO0247
zmo0247_mRNAFw	gtgctgttgctgtcgctatg	RT-PCR	
zmo0247_mRNARv	cgtcaagatgcccatctaaaaa	RT-PCR	
zmo0309_5U_F1	gcctaataggacttttacgacctg	RT-PCR	ZMO0309
zmo0309_mRNAFw	gccaatcgggataaagaacc	RT-PCR	
zmo0309_mRNARv	ggcaaaaaccttatgaccttc	RT-PCR	
zmo0512_5U_F1	ttacaaaggaaagagaacttaactca	RT-PCR	ZMO0512
zmo0512_mRNAFw	ggtggtatctgcttgaattgg	RT-PCR	
zmo0512_mRNARv	gcttgccatcagaatgaaca	RT-PCR	
zmo0120_5U_F1	atcatttgggaggccagag	RT-PCR	ZMO0120
zmo0120_mRNAFw	tcgcagatattgtcgaaaagc	RT-PCR	

zmo0120_mRNARv	cttggaaactgccgatcat	RT-PCR	
zmo0187_5U_F1	gcgcagcttgaaaggaataa	RT-PCR	ZMO0187
zmo0187_mRNAFw	gcgcgacaattaccgattta	RT-PCR	
zmo0187_mRNARv	aacagacgatctgggtcagg	RT-PCR	
zmo0215_5U_F1	tgaatatttttctgtgcttggatg	RT-PCR	ZMO0215
zmo0215_mRNAFw	tattcagccaacagcgtcat	RT-PCR	
zmo0215_mRNARv	aaaaccgactaggggcgtta	RT-PCR	
zmo0369_5U_F1	cctgttgggtagccttctga	RT-PCR	ZMO0369
zmo0369_mRNAFw	tggtcgggttctttctcttg	RT-PCR	
zmo0369_mRNARv	cgaaataacggccttcagtc	RT-PCR	
zmo1181_5U_F1	tttgtgtttttcttaagcgtaaattg	RT-PCR	ZMO1181
zmo1181_mRNAFw	tgatttgggaatcgacaaaag	RT-PCR	
zmo1181_mRNARv	ttaggtcgcttttcaagcaag	RT-PCR	
zmo1682_5U_F1	tcgctttttcctagatatttcaattag	RT-PCR	ZMO1682
zmo1682_mRNAFw	aaacgcaggtcgtggtaatc	RT-PCR	
zmo1682_mRNARv	cgaccatttcccgtaagaca	RT-PCR	
zmo0689_5U_F1	accacgacggatcttctctg	RT-PCR	ZMO0689
zmo0689_mRNAFw	tttcagcaacggatgacaac	RT-PCR	
zmo0689_mRNARv	cagcgtcgatttttggatct	RT-PCR	
zmo1236_5U_F1	ccctatattcgcaagatgtatgtc	RT-PCR	ZMO1236
zmo1236_mRNAFw	ttacgccctctgaaatacgg	RT-PCR	
zmo1236_mRNARv	caagaccatctggcactttg	RT-PCR	
zmo0976_5U_F1	tttaaaaaatgtggagagtttctg	RT-PCR	ZMO0976
zmo0976_mRNAFw	gggcgattgatcttggtatc	RT-PCR	
zmo0976_mRNARv	gcatagggcaggatgtcttt	RT-PCR	
zmo1522_5U_F1	tttggggacggtgcatatct	RT-PCR	ZMO1522
zmo1522_mRNAFw	gcctgtttttgcacaggaa	RT-PCR	
zmo1522_mRNARv	gcatcaaacgacgaccatta	RT-PCR	
zmo0130_5U_F1	tcttatataacaaggaaagactgacga	RT-PCR	ZMO0130
zmo0130_mRNAFw	tgaaccctcctcgacttatca	RT-PCR	
zmo0130_mRNARv	agccttgttttcccagacct	RT-PCR	
zmo1289_5U_F1	ggcgttttatacacaaatattgaca	RT-PCR	ZMO1289
zmo1289_mRNAFw	gatggctcgctggacttatt	RT-PCR	
zmo1289_mRNARv	tcagttgcgtgatttaaagaaa	RT-PCR	
zmo0131_5U_F1	aggttcacgaggaaataggg	RT-PCR	ZMO0131
zmo0131_mRNAFw	aacagcacctcgcctttatg	RT-PCR	

zmo0131_mRNARv	tccagaataacgaaatgcaca	RT-PCR	
zmo0748_5U_F1	cctagacaatcatttttaaggaca	RT-PCR	ZMO0748
zmo0748_mRNAFw	cggctcaatcaaagaccgta	RT-PCR	
zmo0748_mRNARv	ctgttgtggtggcatggata	RT-PCR	
zmo0918_5U_F1	caatcaggtaagagggacgtt	RT-PCR	ZMO0918
zmo0918_mRNAFw	ccaaccctcctccaagatgt	RT-PCR	
zmo0918_mRNARv	ccatcttggggtcttttaaattc	RT-PCR	
zmo1029_5U_F1	gccaataaccctgacgtctt	RT-PCR	ZMO1029
zmo1029_mRNAFw	gcaataacggtcagggaaaa	RT-PCR	
zmo1029_mRNARv	atgtgccaaagccaatcttt	RT-PCR	
zmo1410_5U_F1	agggttgtaaaaagcaaaagg	RT-PCR	ZMO1410
zmo1410_mRNAFw	gggccataaagaggtcattg	RT-PCR	
zmo1410_mRNARv	ccgtctggctgttaagctg	RT-PCR	
zmo1034_5U_F1	cagacaatattcataaactgcctga	RT-PCR	ZMO1034
zmo1034_mRNAFw	gtctgtttttagcggcaacc	RT-PCR	
zmo1034_mRNARv	tgatgatgaccacctttaccg	RT-PCR	
zmo1113_5U_F1	caaaccaccacgccaaag	RT-PCR	ZMO1113
zmo1113_mRNAFw	gaacgtttcgggaaagagg	RT-PCR	
zmo1113_mRNARv	gcggctatccaaaaacagac	RT-PCR	
zmo0753_5U_F1	cagtaacatttagtcggttttctcc	RT-PCR	ZMO0753
zmo0753_mRNAFw	tattgcaaacgggcaaaag	RT-PCR	
zmo0753_mRNARv	ccactttgcgactgataccat	RT-PCR	
zmo1639_5U_F1	tgagcggctttcttcacttt	RT-PCR	ZMO1639
zmo1639_mRNAFw	agccggatactcgggtttat	RT-PCR	
zmo1639_mRNARv	aacagccaactgaaattcacc	RT-PCR	
zmo1350_5U_F1	aatatctgccgcatcagagg	RT-PCR	ZMO1350
zmo1350_mRNAFw	gtctggacgcaccgaagat	RT-PCR	
zmo1350_mRNARv	acaccggccatgataaaaat	RT-PCR	
zmo1576_5U_F1	cggtgcgaagaattttgtct	RT-PCR	ZMO1576
zmo1576_mRNAFw	gtcagatttgggggcaact	RT-PCR	
zmo1576_mRNARv	ggcaaccgaagagacattga	RT-PCR	
zmo1705_5U_F1	taaatatcaaagaaacgggcaat	RT-PCR	ZMO1705
zmo1705_mRNAFw	gctgtatttttcggccacat	RT-PCR	
zmo1705_mRNARv	cccaaaagagtaacggcttg	RT-PCR	
zmo1385_5U_F1	ccggaaataaaaggccagat	RT-PCR	ZMO1385
zmo1385_mRNAFw	cagcttgaagccaaggtgag	RT-PCR	

zmo1385_mRNARv	ataatcgcggagcttgtttc	RT-PCR	
zmo0347_5U_F1	tcttgagagactcttcacataggc	RT-PCR	ZMO0347
zmo0347_mRNAFw	gccgaaaaggtcaacaatct	RT-PCR	
zmo0347_mRNARv	agctgcgcgtcgtcatta	RT-PCR	
zmo1756_5U_F1	aatgctggctttcccatagat	RT-PCR	ZMO1756
zmo1756_mRNAFw	tgattgcaaagtggcgtcta	RT-PCR	
zmo1756_mRNARv	gctggcacgtcctaaggtaa	RT-PCR	
zmo1222_5U_F1	cagcccctcttctggttcta	RT-PCR	ZMO1222
zmo1222_mRNAFw	ctgttgtaacgggtgcctct	RT-PCR	
zmo1222_mRNARv	gcagcataattggcttgtcc	RT-PCR	
zmo1198_5U_F1	tgaaatcgaccagattattttgttt	RT-PCR	ZMO1198
zmo1198_mRNAFw	ggccgctatcgtgtgtttat	RT-PCR	
zmo1198_mRNARv	tggttcaattcatcggaaaa	RT-PCR	
zmo1478_5U_F1	aaaaggcagaaatgccgata	RT-PCR	ZMO1478
zmo1478_mRNAFw	tcaagcaagccattgagaaa	RT-PCR	
zmo1478_mRNARv	aaattgggacctgggaaaat	RT-PCR	
zmo0275_5U_F1	tttccccgatatggatttga	RT-PCR	ZMO0275
zmo0275_mRNAFw	atcgcgaaatcctacgatca	RT-PCR	
zmo0275_mRNARv	gtctttcaccaccggaaagc	RT-PCR	
zmo0778_5U-_F1	tccgtatcaacgatctggaa	RT-PCR	ZMO0778
zmo0778_mRNAFw	aaatcccgattgtggcagt	RT-PCR	
zmo0778_mRNARv	aaggtcattcagctgctgttc	RT-PCR	
zmo1673_5U_F1	cagcaagaaaggattggagtc	RT-PCR	ZMO1673
zmo1673_mRNAFw	tggtcgctttgggatatgat	RT-PCR	
zmo1673_mRNARv	tttttgctggaaataagggtgt	RT-PCR	
zmo1497_5U_F1	tcttatcgggggataatga	RT-PCR	ZMO1497
zmo1497_mRNAFw	ttttgtgttaacggcggaag	RT-PCR	
zmo1497_mRNARv	gcagcgggaaggttgataat	RT-PCR	
zmo1770_5U_F1	gtaatgacccagtcgggaga	RT-PCR	ZMO1770
zmo1770_mRNAFw	atattgccggttccaaagc	RT-PCR	
zmo1770_mRNARv	gtaaagcccggcatcacg	RT-PCR	
zmo1107_5U_F1	tatctagccggaagattaaatagaagg	RT-PCR	ZMO1107
zmo1107_mRNAFw	tgacttggcacgtaaagcag	RT-PCR	
zmo1107_mRNARv	attactctttggtatcaacaggaaca	RT-PCR	
zmo1407_5U_F1	tatatcggtcattccatccttg	RT-PCR	ZMO1407
zmo1407_mRNAFw	gtgaaatgttggcgattctg	RT-PCR	
zmo1407_mRNARv	caacaacgcgtttgatcttt	RT-PCR	

zmo1033_5U_F1	caagttaaagcatgtaaactctgacc	RT-PCR	ZMO1033
zmo1033_mRNAFw	ttatatgggatcgggtgagg	RT-PCR	
zmo1033_mRNARv	ccagcttggatgcgatgt	RT-PCR	
zmo1008_5U_F1	gtctatgcgagcttcggttc	RT-PCR	ZMO1008
zmo1008_mRNAFw	atgatttgcgccagctattg	RT-PCR	
zmo1008_mRNARv	tctttcttttcagtggcatcg	RT-PCR	
zmo1399_5U_F1	acttcggttccgattgtgac	RT-PCR	ZMO1399
zmo1399_mRNAFw	ttggacggatgacattgaaa	RT-PCR	
zmo1399_mRNARv	acctccgaaacccccataa	RT-PCR	
zmo0179_5U_F1	ttaccttatatcccaagggaagg	RT-PCR	ZMO0179
zmo0179_mRNAFw	cgggacaaaggctaatcttg	RT-PCR	
zmo0179_mRNARv	cgcaatatggcggaattctt	RT-PCR	
zmo1235_5U_F1	atataatagaaggaaatctggcct	RT-PCR	ZMO1235
zmo1235_mRNAFw	cgtgcctccgtttacaaag	RT-PCR	
zmo1235_mRNARv	aatgtgaacatttaccccgtattt	RT-PCR	
zmo1411_5U_F1	caaaatttgacggagaaatgtt	RT-PCR	ZMO1411
zmo1411_mRNAFw	tctctttacgaacgggcatc	RT-PCR	
zmo1411_mRNARv	ggaatcgtcatcaaaataaccaa	RT-PCR	
zmo1432_5U_F1	tcttgctatgataggcgactg	RT-PCR	ZMO1432
zmo1432_mRNAFw	gggcgatgacgactgtctat	RT-PCR	
zmo1432_mRNARv	tcgcaatcatggagaatctg	RT-PCR	
zmo0367_5U-F1	aaggattcggcctctgtttt	RT-PCR	ZMO0367
zmo0367_mRNAFw	gatctgcgtatcgtctgcac	RT-PCR	
zmo0367_mRNARv	aatatgatcggaagaagcaaga	RT-PCR	
zmo1566_5U_F1	tttggatagagggactatatcttgtg	RT-PCR	ZMO1556
zmo1566_mRNAFw	tggcacagagcatgaaaaac	RT-PCR	
zmo1566_mRNARv	tttaggcattaaaggcaggtc	RT-PCR	
zmo1599_5U_F1	tgtaacagctaaggcgcttg	RT-PCR	ZMO1599
zmo1599_mRNAFw	gctggtctctgcctatttcg	RT-PCR	
zmo1599_mRNARv	atcttgaacatccggcaact	RT-PCR	
zmo1596_5U_F1	gttgttttcgggttgttgct	RT-PCR	ZMO1596
zmo1596_mRNAFw	aaaatgcgctgatcgtttct	RT-PCR	
zmo1596_mRNARv	acgtgacggtcaacaatgg	RT-PCR	
zmo1771_5U_F1	tacgcataagaaggcggaaa	RT-PCR	ZMO1771
zmo1771_mRNAFw	cggttctagcgctcaaaaga	RT-PCR	
zmo1771_mRNARv	attgacttgcgggttaccac	RT-PCR	

zmo1233_5U_F1	ctcggctcccgacaatagta	RT-PCR	ZMO1233
zmo1233_mRNAFw	ggcgaaaatagcctgtttca	RT-PCR	
zmo1233_mRNARv	aaccgtaacctgtttcaggaga	RT-PCR	
zmo1412_5U_F1	tttaaagcttctctggaaagataga	RT-PCR	ZMO1412
zmo1412_mRNAFw	tatcgtttcagcccatgtca	RT-PCR	
zmo1412_mRNARv	gccacatttgtggtaatgctt	RT-PCR	
zmo0178_5U_F1	ggaggctgtctccgttatca	RT-PCR	ZMO0178
zmo0178_mRNAFw	atcgcgttacggatgatacc	RT-PCR	
zmo0178_mRNARv	ccgagaaagcatcattgaca	RT-PCR	
zmo1048_5U_F1	ggaaaaagatgatagctgtcagaa	RT-PCR	ZMO1048
zmo1048_mRNAFw	acgctgacccttgttgctat	RT-PCR	
zmo1048_mRNARv	cccgcaaccactctacctaa	RT-PCR	
zmo0140_5U_F1	aacgatggatgaatccgttta	RT-PCR	ZMO0140
zmo0140_mRNAFw	gaaggggcttttcgtgatct	RT-PCR	
zmo0140_mRNARv	ggttgagacatcttgccatgt	RT-PCR	
zmo1606_5U_F1	tcgatttatcctgtccgtca	RT-PCR	ZMO1606
zmo1606_mRNAFw	cacgatttgccaccgatac	RT-PCR	
zmo1606_mRNARv	ccccaacgataccatttcc	RT-PCR	
zmo0322_5U_F1	cctgtgttcttatgtaaatcgagac	RT-PCR	ZMO0322
zmo0322_mRNAFw	catttaggccatcaggctgt	RT-PCR	
zmo0322_mRNARv	tctcgaatacgggtagaggaaa	RT-PCR	
zmo0366_5U_F1	agattaaggcgggagaggaa	RT-PCR	ZMO0366
zmo0366_mRNAFw	cattttattgcccctcgtca	RT-PCR	
zmo0366_mRNARv	catctgctgaccagaaacca	RT-PCR	
zmo1097_5U_F1	aattcatatccccgtcaggag	RT-PCR	ZMO1097
zmo1097_mRNAFw	cacctgctttgggtgaaatc	RT-PCR	
zmo1097_mRNARv	ttaaagtgaactttcaatccatga	RT-PCR	
zmo0765_5U_F1	catctttgaaataaaagaaagacga	RT-PCR	ZMO0765
zmo0765_mRNAFw	gcccgttttaatggtcagat	RT-PCR	
zmo0765_mRNARv	atttttcgccctgagattcc	RT-PCR	
zmo1612_5U_F1	gccttgaaaaccgacttctc	RT-PCR	ZMO1612
zmo1612_mRNAFw	cattgcgcttgtctcatttg	RT-PCR	
zmo1612_mRNARv	tagccctgactggtgtgttc	RT-PCR	
zmo0054_5U_F1	agacggcttgcgtgagac	RT-PCR	ZMO0054
zmo0054_mRNAFw	gaattggatgcgtcttcgat	RT-PCR	
zmo0054_mRNARv	ctttttcggtaagggcaacc	RT-PCR	

zmo1289_5U_F1	ggcgttttatacacaaatattgaca	RT-PCR	ZMO1289
zmo1289_mRNAFw	gatggctcgctggacttatt	RT-PCR	
zmo1289_mRNARv	tcagttgcgtgatttaaagaaa	RT-PCR	
zmo0031_5U_F1	cccgacagccttccaaag	RT-PCR	ZMO0031
zmo0031_mRNAFw	aatactggccgttccatcag	RT-PCR	
zmo0031_mRNARv	tcaattttggtcgcattgaa	RT-PCR	
zmo0131_5U_F1	aggttcacgaggaaataggg	RT-PCR	ZMO0131
zmo0131_mRNAFw	aacagcacctcgcctttatg	RT-PCR	
zmo0131_mRNARv	tccagaataacgaaatgcaca	RT-PCR	
zmo0748_5U_F1	gcctagacaatcatttttaaggaca	RT-PCR	ZMO0748
zmo0748_mRNAFw	cggctcaatcaaagaccgta	RT-PCR	
zmo0748_mRNARv	ctgttgtggtggcatggata	RT-PCR	
zmo0918_5U_F1	caatcaggtaagagggacgtt	RT-PCR	ZMO0918
zmo0918_mRNAFw	ccaaccctcctccaagatgt	RT-PCR	
zmo0918_mRNARv	ccatcttggggtcttttaaattc	RT-PCR	
zmo1029_5U_F1	gccaataaccctgacgtctt	RT-PCR	ZMO1029
zmo1029_mRNAFw	gcaataacggtcagggaaaa	RT-PCR	
zmo1029_mRNARv	atgtgccaaagccaatcttt	RT-PCR	
zmo1410_5U_F1	agggttgtaaaaagcaaaagg	RT-PCR	ZMO1410
zmo1410_mRNAFw	gggccataaagaggtcattg	RT-PCR	
zmo1410_mRNARv	ccgtctggctgttaagctg	RT-PCR	
zmo1034_5U_F1	cagacaatattcataaactgcctga	RT-PCR	ZMO1034
zmo1034_mRNAFw	gtctgtttttagcggcaacc	RT-PCR	
zmo1034_mRNARv	tgatgatgaccacctttaccg	RT-PCR	
zmo1113_5U_F1	caaaccaccacgccaaag	RT-PCR	ZMO1113
zmo1113_mRNAFw	gaacgtttcgggaaagagg	RT-PCR	
zmo1113_mRNARv	gcggctatccaaaaacagac	RT-PCR	
zmo0753_5U_F1	cagtaacatttagtcggttttctcc	RT-PCR	ZMO0753
zmo0753_mRNAFw	tattgcaaacgggcaaaag	RT-PCR	
zmo0753_mRNARv	ccactttgcgactgataccat	RT-PCR	
zmo1639_5U_F1	tgagcggctttcttcacttt	RT-PCR	ZMO1639
zmo1639_mRNAFw	agccggatactcgggtttat	RT-PCR	
zmo1639_mRNARv	aacagccaactgaaattcacc	RT-PCR	
zmo1350_5U_F1	aatatctgccgcatcagagg	RT-PCR	ZMO1350
zmo1350_mRNAFw	gtctggacgcaccgaagat	RT-PCR	
zmo1350_mRNARv	acaccggccatgataaaaat	RT-PCR	
zmo1576_5U_F1	cggtgcgaagaattttgtct	RT-PCR	ZMO1576

zmo1576_mRNAFw	gtcagatttgggggcaact	RT-PCR	
zmo1576_mRNARv	ggcaaccgaagagacattga	RT-PCR	
zmo1705_5U_F1	taaatatcaaagaaacgggcaat	RT-PCR	ZMO1705
zmo1705_mRNAFw	gctgtatttttcggccacat	RT-PCR	
zmo1705_mRNARv	cccaaaagagtaacggcttg	RT-PCR	
zmo1385_5U_F1	ccggaaataaaaggccagat	RT-PCR	ZMO1385
zmo1385_mRNAFw	cagcttgaagccaaggtgag	RT-PCR	
zmo1385_mRNARv	ataatcgcggagcttgtttc	RT-PCR	
zmo0347_5U_F1	tcttgagagactcttcacataggc	RT-PCR	ZMO0347
zmo0347_mRNAFw	gccgaaaaggtcaacaatct	RT-PCR	
zmo0347_mRNARv	agctgcgcgtcgtcatta	RT-PCR	
zmo1756_5U_F1	aatgctggctttcccatagat	RT-PCR	ZMO1756
zmo1756_mRNAFw	tgattgcaaagtggcgtcta	RT-PCR	
zmo1756_mRNARv	gctggcacgtcctaaggtaa	RT-PCR	
zmo1222_5U_F1	cagcccctcttctggttcta	RT-PCR	ZMO1222
zmo1222_mRNAFw	ctgttgtaacgggtgcctct	RT-PCR	
zmo1222_mRNARv	gcagcataattggcttgtcc	RT-PCR	
zmo1198_5U_F1	tgaaatcgaccagattattttgttt	RT-PCR	ZMO1198
zmo1198_mRNAFw	ggccgctatcgtgtgtttat	RT-PCR	
zmo1198_mRNARv	tggttcaattcatcggaaaa	RT-PCR	
zmo1478_5U_F1	aaaaggcagaaatgccgata	RT-PCR	ZMO1478
zmo1478_mRNAFw	tcaagcaagccattgagaaa	RT-PCR	
zmo1478_mRNARv	aaattgggacctgggaaaat	RT-PCR	
zmo0275_5U_F1	tttccccgatatggatttga	RT-PCR	ZMO0275
zmo0275_mRNAFw	atcgcgaaatcctacgatca	RT-PCR	
zmo0275_mRNARv	gtctttcaccaccggaaagc	RT-PCR	
zmo0778_5U-_F1	tccgtatcaacgatctggaa	RT-PCR	ZMO0778
zmo0778_mRNAFw	aaatcccgattgtggcagt	RT-PCR	
zmo0778_mRNARv	aaggtcattcagctgctgttc	RT-PCR	
zmo1673_5U_F1	cagcaagaaaggattggagtc	RT-PCR	ZMO1673
zmo1673_mRNAFw	tggtcgctttgggatatgat	RT-PCR	
zmo1673_mRNARv	tttttgctggaaataagggtgt	RT-PCR	
zmo1497_5U_F1	tcttatcgggggataatga	RT-PCR	ZMO1497
zmo1497_mRNAFw	ttttgtgttaacggcggaag	RT-PCR	
zmo1497_mRNARv	gcagcgggaaggttgataat	RT-PCR	
zmo1770_5U_F1	gtaatgacccagtcgggaga	RT-PCR	ZMO1770
zmo1770_mRNAFw	atattgccggttccaaagc	RT-PCR	

zmo1770_mRNARv	gtaaagcccggcatcacg	RT-PCR	
zmo1107_5U_F1	tatctagccggaagattaaatagaagg	RT-PCR	ZMO1107
zmo1107_mRNAFw	tgacttggcacgtaaagcag	RT-PCR	
zmo1107_mRNARv	attactctttggtatcaacaggaaca	RT-PCR	
zmo1407_5U_F1	tatatcggtcattccatccttg	RT-PCR	ZMO1407
zmo1407_mRNAFw	gtgaaatgttggcgattctg	RT-PCR	
zmo1407_mRNARv	caacaacgcgtttgatcttt	RT-PCR	
zmo1033_5U_F1	caagttaaagcatgtaaactctgacc	RT-PCR	ZMO1033
zmo1033_mRNAFw	ttatatgggatcgggtgagg	RT-PCR	
zmo1033_mRNARv	ccagcttggatgcgatgt	RT-PCR	
zmo1008_5U_F1	gtctatgcgagcttcggttc	RT-PCR	ZMO1008
zmo1008_mRNAFw	atgatttgcgccagctattg	RT-PCR	
zmo1008_mRNARv	tctttcttttcagtggcatcg	RT-PCR	
zmo1399_5U_F1	acttcggttccgattgtgac	RT-PCR	ZMO1399
zmo1399_mRNAFw	ttggacggatgacattgaaa	RT-PCR	
zmo1399_mRNARv	acctccgaaacccccctataa	RT-PCR	
zmo0179_5U_F1	ttaccttatatcccaagggaagg	RT-PCR	ZMO0179
zmo0179_mRNAFw	cgggacaaaggctaatcttg	RT-PCR	
zmo0179_mRNARv	cgcaatatggcggaattctt	RT-PCR	
zmo1235_5U_F1	atataatagaaggaaatctggcct	RT-PCR	ZMO1235
zmo1235_mRNAFw	cgtgcctccgtttacaaag	RT-PCR	
zmo1235_mRNARv	aatgtgaacatttaccccgtattt	RT-PCR	
zmo1411_5U_F1	caaaatttgacggagaaatgtt	RT-PCR	ZMO1411
zmo1411_mRNAFw	tctctttacgaacgggcatc	RT-PCR	
zmo1411_mRNARv	ggaatcgtcatcaaaataaccaa	RT-PCR	
zmo1432_5U_F1	tcttgctatgataggcgactg	RT-PCR	ZMO1432
zmo1432_mRNAFw	gggcgatgacgactgtctat	RT-PCR	
zmo1432_mRNARv	tcgcaatcatggagaatctg	RT-PCR	
zmo0367_5U-F1	aaggattcggcctctgtttt	RT-PCR	ZMO0367
zmo0367_mRNAFw	gatctgcgtatcgtctgcac	RT-PCR	
zmo0367_mRNARv	aatatgatcggaagaagcaaga	RT-PCR	
zmo1566_5U_F1	tttggatagagggactatatcttgtg	RT-PCR	ZMO1556
zmo1566_mRNAFw	tggcacagagcatgaaaaac	RT-PCR	
zmo1566_mRNARv	tttaggcattaaaggcaggtc	RT-PCR	
zmo1599_5U_F1	tgtaacagctaaggcgcttg	RT-PCR	ZMO1599
zmo1599_mRNAFw	gctggtctctgcctatttcg	RT-PCR	
zmo1599_mRNARv	atcttgaacatccggcaact	RT-PCR	

zmo1596_5U_F1	gttgttttcgggttgttgct	RT-PCR	ZMO1596
zmo1596_mRNAFw	aaaatgcgctgatcgtttct	RT-PCR	
zmo1596_mRNARv	acgtgacggtcaacaatgg	RT-PCR	
zmo1771_5U_F1	tacgcataagaaggcggaaa	RT-PCR	ZMO1771
zmo1771_mRNAFw	cggttctagcgctcaaaaga	RT-PCR	
zmo1771_mRNARv	attgacttgcgggttaccac	RT-PCR	
zmo1233_5U_F1	ctcggctcccgacaatagta	RT-PCR	ZMO1233
zmo1233_mRNAFw	ggcgaaaatagcctgtttca	RT-PCR	
zmo1233_mRNARv	aaccgtaacctgtttcaggaga	RT-PCR	
zmo1412_5U_F1	tttaaagcttctctggaaagataga	RT-PCR	ZMO1412
zmo1412_mRNAFw	tatcgtttcagcccatgtca	RT-PCR	
zmo1412_mRNARv	gccacatttgtggtaatgctt	RT-PCR	
zmo0178_5U_F1	ggaggctgtctccgttatca	RT-PCR	ZMO0178
zmo0178_mRNAFw	atcgcgttacggatgatacc	RT-PCR	
zmo0178_mRNARv	ccgagaaagcatcattgaca	RT-PCR	
zmo1048_5U_F1	ggaaaaagatgatagctgtcagaa	RT-PCR	ZMO1048
zmo1048_mRNAFw	acgctgacccttgttgctat	RT-PCR	
zmo1048_mRNARv	cccgcaaccactctacctaa	RT-PCR	
zmo0140_5U_F1	aacgatggatgaatccgttta	RT-PCR	ZMO0140
zmo0140_mRNAFw	gaaggggcttttcgtgatct	RT-PCR	
zmo0140_mRNARv	ggttgagacatcttgccatgt	RT-PCR	
zmo1606_5U_F1	tcgatttatcctgtccgtca	RT-PCR	ZMO1606
zmo1606_mRNAFw	cacgatttgccaccgatac	RT-PCR	
zmo1606_mRNARv	ccccaacgataccatttcc	RT-PCR	
zmo0322_5U_F1	cctgtgttcttatgtaaatcgagac	RT-PCR	ZMO0322
zmo0322_mRNAFw	catttaggccatcaggctgt	RT-PCR	
zmo0322_mRNARv	tctcgaatacgggtagaggaaa	RT-PCR	
zmo0366_5U_F1	agattaaggcgggagaggaa	RT-PCR	ZMO0366
zmo0366_mRNAFw	cattttattgcccctcgtca	RT-PCR	
zmo0366_mRNARv	catctgctgaccagaaacca	RT-PCR	
zmo1097_5U_F1	aattcatatccccgtcaggag	RT-PCR	ZMO1097
zmo1097_mRNAFw	cacctgctttgggtgaaatc	RT-PCR	
zmo1097_mRNARv	ttaaagtgaactttcaatccatga	RT-PCR	
zmo0765_5U_F1	catctttgaaataaaagaaagacga	RT-PCR	ZMO0765
zmo0765_mRNAFw	gcccgttttaatggtcagat	RT-PCR	
zmo0765_mRNARv	atttttcgccctgagattcc	RT-PCR	
zmo1612_5U_F1	gccttgaaaaccgacttctc	RT-PCR	ZMO1612

zmo1612_mRNAFw	cattgcgcttgtctcatttg	RT-PCR	
zmo1612_mRNARv	tagccctgactggtgtgttc	RT-PCR	
zmo0054_5U_F1	agacggcttgcgtgagac	RT-PCR	ZMO0054
zmo0054_mRNAFw	gaattggatgcgtcttcgat	RT-PCR	
zmo0054_mRNARv	ctttttcggtaagggcaacc	RT-PCR	
5RACE_RSE1	TGACATTGGCACCCGCTTTG	5' RACE	
5RACE_RSE2	GAAATCGCCATGCCCTGACTG	5' RACE	
5RACE_RSE3	TTGGTGTCGAACCATTTGGTC	5' RACE	
5RACE_RSE5	CATCAAGAAGATGGGCGACAACC	5' RACE	
5RACE_RSE6	TTTCGACGGAGGTATCCTCAAG	5' RACE	
5RACE_RSE7	CATTAACAAATCCCGTCCGCACTG	5' RACE	
5RACE_RSE8	CTGAAGGATGAAAGCCGAAATCTGG	5' RACE	
5RACE_RSE9	ATGCGGCTTGGCAGTGATTTC	5' RACE	
5RACE_RSE10	TTTCTTCCGCTTGACGTTTGG	5' RACE	
5RACE_RSE11	CTTGGTGCAATGGCGGGTAATG	5' RACE	
5RACE_RSE12	GAAACCATCACAGGTAGCGCAAG	5' RACE	
5RACE_RSE13	TCTACAGAATGGGCAGGATCAGG	5' RACE	
5RACE_RSE14	GAAGCAAAGCATCCCTGTGG	5' RACE	
5RACE_RSE15	CGTCTTTCGGGAAGCGCATC	5' RACE	
5RACE_RSE16	TGCCAAGAGCGTGATGCAAC	5' RACE	
5RACE_RSE17	TTTGGCTAATTCGGGATCGAGATG	5' RACE	
5RACE_RSE18	ATTGCGCGGATTGTGATTTCTTCG	5' RACE	
5RACE_RSE19	AGCTTGGCAAAGAAACGCCATC	5' RACE	

5RACE_RSE20	TCCCGGACCAAGAATAGTGATAAC	5' RACE
5RACE_RSE21	TGACAGCATCCAATCGGCTC	5' RACE
5RACE_RSE22	CGGATTGCCTGCATGGATGAG	5' RACE
5RACE_RSE23	TCTGTGCTAAAGATCGTGGAG	5' RACE
5RACE_RSE24	CCATAAAGGACGATATTCGCACC	5' RACE
5RACE_RSE26	CATCCCAAGATGCCTTCAACC	5' RACE
5RACE_RSE27	GGGATCGACGCCATATTCAGC	5' RACE
5RACE_RSE28	GATCATCGGCGGTGCGTTTG	5' RACE
5RACE_RSE29	GGGAATCCAAGCGCCTTCTG	5' RACE
5RACE_RSE30	TGTTCCCGACTAAATCCCAATTCC	5' RACE
5RACE_RSE31	AAAGCCAATCTTTGCCGCATC	5' RACE
5RACE_RSE32	ACCACCTTTACCGTCTTCTGGAG	5' RACE
5RACE_RSE33	CACCTGGCGTATTGAAATCATTGG	5' RACE
5RACE_RSE34	CATTACGGGTCAACAAGACGCAG	5' RACE
5RACE_RSE35	AATCGTGGACAAGGTCGCTTC	5' RACE
5RACE_RSE36	GCCATCATTACCCATACCCAACC	5' RACE
5RACE_RSE37	TGGAGTTCAAGGACGGAATCG	5' RACE
5RACE_RSE38	GGATGACCAAAGCCGTCTGAAG	5' RACE
5RACE_RSE39	CGCTGGAAACGATTGGTCAGATG	5' RACE
5RACE_RSE40	CATGAAGCGCGGACAAGATCAC	5' RACE
5RACE_RSE41	GCCGACAGCATAACCGATACAG	5' RACE

5RACE_RSE42	CTACGCTAAGAAAGCCGATAACAG	5' RACE	
5RACE_RSE43	AGAAGTTGGACCAGCCAGACC	5' RACE	
5RACE_RSE44	GCATTAAAGGCAGGTCGCTACG	5' RACE	
5RACE_RSE45	CGGTATTGTTCCGGCGTCAG	5' RACE	
5RACE_RSE46	AATACCGGCCAACAATTCAACC	5' RACE	
5RACE_RSE47	GGTCAGCTACCGATTTGCCTTC	5' RACE	
5RACE_RSE48	CACCGATACCTAAACCGGCAAG	5' RACE	
5RACE_RSE49	AGCAGCAGGGTCGGGATATTG	5' RACE	
5RACE_RSE50	ACAATACCGCTGCACCATCC	5' RACE	
GGRSEZMO0172F	agatt CGTCTC TCTCCC tccccgggggccgtataa	golden gate cloning of UTR+90bps	ZMO0172
GGRSEZMO0172R	agatt CGTCTC CACCAT atttttaccttcaacgtaaattttacgggaa	golden gate cloning of UTR+90bps	
GGRSEZMO0979F	agatt CGTCTC TCTCCC ggaaatttttttgcatagg	golden gate cloning of UTR+90bps	ZMO0979
GGRSEZMO0979R	agatt CGTCTC CACCAT ctgataggcttctttg	golden gate cloning of UTR+90bps	
GGRSEZMO1000F	agatt CGTCTC TCTCCC aaggattaaggttctttgtc	golden gate cloning of UTR+90bps	ZMO1000
GGRSEZMO1000R	agatt CGTCTC CACCAT acggccaccccagaaac	golden gate cloning of UTR+90bps	
GGRSEZMO0547F	agatt CGTCTC TCTCCC gaatgaccatttccatt	golden gate cloning of UTR+90bps	ZMO0547
GGRSEZMO0547R	agatt CGTCTC CACCAT aagccgataaggtttgtat	golden gate cloning of UTR+90bps	
GGRSEZMO0056F	agatt CGTCTC TCTCCC atctcaaagacagtccg	golden gate cloning of UTR+90bps	ZMO0056
GGRSEZMO0056R	agattCGTCTCCACCATatcata accgcgatattcc	golden gate cloning of UTR+90bps	
GGRSEZMO0376F	agatt CGTCTC TCTCCC aaaaggtgtgagcattttg	golden gate cloning of UTR+90bps	ZMO0376
GGRSEZMO0376R	agatt CGTCTC CACCAT cgatttttcacgaccga	golden gate cloning of UTR+90bps	

GGRSEZMO0546F	agatt CGTCTC TCTCCC aattttggtcacattgt	golden gate cloning of UTR+90bps	ZMO0546
GGRSEZMO0546R	agatt CGTCTC CACCAT aataagcgcgaaggtcg	golden gate cloning of UTR+90bps	
GGRSEZMO0660F	agatt CGTCTC TCTCCC gtagacttgttccaaggg	golden gate cloning of UTR+90bps	ZMO0660
GGRSEZMO0660R	agatt CGTCTC CACCAT tgcgttttcaataacctt	golden gate cloning of UTR+90bps	
GGRSEZMO1069F	agatt CGTCTC TCTCCC tttgctggtggacagaa	golden gate cloning of UTR+90bps	ZMO1069
GGRSEZMO1069R	agatt CGTCTC CACCAT taatcttactttctgggcagc	golden gate cloning of UTR+90bps	
GGRSEZMO1139F	agatt CGTCTC TCTCCC cccttatcaaggctacaa	golden gate cloning of UTR+90bps	ZMO1139
GGRSEZMO1139R	agatt CGTCTC CACCAT tgctccgccgggatagc	golden gate cloning of UTR+90bps	
GGRSEZMO1142F	agatt CGTCTC TCTCCC tctgctgtagagtgaca	golden gate cloning of UTR+90bps	ZMO1142
GGRSEZMO1142R	agatt CGTCTC CACCAT tccggcacgagccg	golden gate cloning of UTR+90bps	
GGRSEZMO0709F	agatt CGTCTC TCTCCC acatgcaacctttgaa	golden gate cloning of UTR+90bps	ZMO0709
GGRSEZMO0709R	agatt CGTCTC CACCAT gggtgcaatggctttga	golden gate cloning of UTR+90bps	
GGRSEZMO1137F	agatt CGTCTC TCTCCC ggagaactggatgtttt	golden gate cloning of UTR+90bps	ZMO1137
GGRSEZMO1137R	agatt CGTCTC CACCAT tttaccgccctttgaag	golden gate cloning of UTR+90bps	
GGRSEZMO0187F	agatt CGTCTC TCTCCC ttttttgccagagcttt	golden gate cloning of UTR+90bps	ZMO0187
GGRSEZMO0187R	agatt CGTCTC CACCAT agccttatccagtttatcg	golden gate cloning of UTR+90bps	
GGRSEZMO0369F	agatt CGTCTC TCTCCC tttacctgttgggtagc	golden gate cloning of UTR+90bps	ZMO0369
GGRSEZMO0369R	agatt CGTCTC CACCAT tccaagagaaagaaccc	golden gate cloning of UTR+90bps	
GGRSEZMO0405F	agatt CGTCTC TCTCCC agacgagcttatttagg	golden gate cloning of UTR+90bps	ZMO0405
GGRSEZMO0405R	agatt CGTCTC CACCAT gacggctggcgatag	golden gate cloning of UTR+90bps	
GGRSEZMO0937F	agatt CGTCTC TCTCCC aatatccctctttctat	golden gate cloning of UTR+90bps	ZMO0937

GGRSEZMO0937R	agatt CGTCTC CACCAT tgttaccgatgaaagac	golden gate cloning of UTR+90bps	
GGRSEZMO1179F	agatt CGTCTC TCTCCC gtaacagtcgcctcatg	golden gate cloning of UTR+90bps	ZMO1179
GGRSEZMO1179R	agatt CGTCTC CACCAT tatgatgcgccccgaat	golden gate cloning of UTR+90bps	
GGRSE24F	agatt CGTCTC TCTCCC aatactatctgttttag att	golden gate cloning of UTR+90bps	ZMO1275
GGRSE24R	agatt CGTCTC CACCAT tctaataatggcattct	golden gate cloning of UTR+90bps	
GGRSE27F	agatt CGTCTC TCTCCC ctggaaaagccggcgaat	golden gate cloning of UTR+90bps	ZMO0689
GGRSE27R	agatt CGTCTC CACCAT acgggtcagatttggcgt	golden gate cloning of UTR+90bps	
GGRSE28F	agatt CGTCTC TCTCCC ctggcctcatttctctc	golden gate cloning of UTR+90bps	ZMO0131
GGRSE28R	agatt CGTCTC CACCAT aacaccggcataaaggc	golden gate cloning of UTR+90bps	
GGRSE29F	agatt CGTCTC TCTCCC aagacgcgccctacata	golden gate cloning of UTR+90bps	ZMO0748
GGRSE29R	agatt CGTCTC CACCAT tatccatacttcatggt	golden gate cloning of UTR+90bps	
GGRSE32F	agatt CGTCTC TCTCCC aatcagacaatattcat	golden gate cloning of UTR+90bps	ZMO1034
GGRSE32R	agatt CGTCTC CACCAT ttgggcggaagcct	golden gate cloning of UTR+90bps	
GGRSE33F	agatt CGTCTC TCTCCC aattgaggcggctcgca	golden gate cloning of UTR+90bps	ZMO1113
GGRSE33R	agatt CGTCTC CACCAT gaaacgttcccccaaag	golden gate cloning of UTR+90bps	
GGRSE34F	agatt CGTCTC TCTCCC atcccgatggcgtttttc	golden gate cloning of UTR+90bps	ZMO0347
GGRSE34R	agatt CGTCTC CACCAT ctgtgactgaccatctc	golden gate cloning of UTR+90bps	
GGRSE35F	agatt CGTCTC TCTCCC attcttccttcttatag	golden gate cloning of UTR+90bps	ZMO1198
GGRSE35R	agatt CGTCTC CACCAT tttattccgcaaaatat	golden gate cloning of UTR+90bps	
GGRSE36F	agatt CGTCTC TCTCCC atattctcaatagtctg	golden gate cloning of UTR+90bps	ZMO1478
GGRSE36R	agatt CGTCTC CACCAT ggcttgcttgataataa	golden gate cloning of UTR+90bps	

GGRSE37F	agatt CGTCTC TCTCCC ataaaaagcctatcttg	golden gate cloning of UTR+90bps	ZMO0275
GGRSE37R	agatt CGTCTC CACCAT cgcgatagaaataatgg	golden gate cloning of UTR+90bps	
GGRSE41F	agatt CGTCTC TCTCCC aataaaagcaacaacctt	golden gate cloning of UTR+90bps	ZMO1399
GGRSE41R	agatt CGTCTC CACCAT caactctttcaatgtca	golden gate cloning of UTR+90bps	
GGRSE42F	agatt CGTCTC TCTCCC ttttacaaataggcataa	golden gate cloning of UTR+90bps	ZMO1432
GGRSE42R	agatt CGTCTC CACCAT aagcccaatccgcatag	golden gate cloning of UTR+90bps	
GGRSE43F	agatt CGTCTC TCTCCC ttaaacttgctttggctg	golden gate cloning of UTR+90bps	ZMO0367
GGRSE43R	agatt CGTCTC CACCAT atcaagaccataaagcg	golden gate cloning of UTR+90bps	
GGRSE45F	agatt CGTCTC TCTCCC gtaaacgttgggagatc	golden gate cloning of UTR+90bps	ZMO1412
GGRSE45R	agatt CGTCTC CACCAT aacagaaacgctgttgt	golden gate cloning of UTR+90bps	
GGRSE46F	agatt CGTCTC TCTCCC gtatattttggctgtttg	golden gate cloning of UTR+90bps	ZMO1048
GGRSE46R	agatt CGTCTC CACCAT gcacagggaacggaaaa	golden gate cloning of UTR+90bps	
GGRSE47F	agatt CGTCTC TCTCCC agcgctacagaaaaata	golden gate cloning of UTR+90bps	ZMO0140
GGRSE47R	agatt CGTCTC CACCAT gctgggcgagatcac	golden gate cloning of UTR+90bps	
GGRSE48F	agatt CGTCTC TCTCCC agtttatcgccaaggat	golden gate cloning of UTR+90bps	ZMO0366
GGRSE48R	agatt CGTCTC CACCAT aaccgctgaatcgtaac	golden gate cloning of UTR+90bps	
GGRSE50F	agatt CGTCTC TCTCCC attgattcgcccgaata	golden gate cloning of UTR+90bps	ZMO1612
GGRSE50R	agatt CGTCTC CACCAT tggggtctggttatcaat	golden gate cloning of UTR+90bps	
DelUTRupF	ATGTGAATTCGGTTTGGT CTTCTGACGGTTATTTC	Forward primer for up homology arm (EcoRI)	ZMO0347
DelUTRupR	ATGTGGTACCCGCCGCTG GACCTACCAAGGCAACGC TATGTTCTCTTG	Reverse primer for up homology arm (KpnI)	

DelUTRdownF	ATGAGTCGACCAAGAGAA CATAGCGTTGCCTTGGTA GGTCCAGCGGCG	Forward primer for down homology arm (SalI)
DelUTRdownR	ATGTGCATGCGGATCATG CGGCGATCGGCTTCAAT	Reverse primer for down homology arm (SphI)

Table A.4: Transcripts significantly dependent on Zms4 and Zms6

Values from Likelihood Ratio Test ['~ strain + time + strain:time' vs '~ time'] calculated by DESeq2. Biological triplicates. $p_{adj} < 0.01$

GeneID	log2(pZms4/wt)	p_{adj}
ZMO1335	4.83	5.20E-110
Zms4	7.98	9.87E-67
Zms6	-1.33	9.94E-35
ZMO0204	-0.32	4.59E-26
ZMO0989	-0.74	1.16E-24
ZMO1025	0.35	1.05E-20
ZMO1941	2.86	1.56E-20
ZMO1928	0.02	7.89E-17
ZMO1929	-0.07	4.51E-16
sRNA100	-0.32	8.52E-16
ZMO2037	4.15	1.49E-15
ZMO1754	-0.99	6.43E-15
ZMO1424	0.02	4.89E-14
ZMO1721	-0.17	3.45E-13
ZMO1485	-0.39	1.94E-12
ZMO0725	0.14	2.11E-12
ZMO0256	0.21	2.11E-12
ZMO1373	-0.88	2.96E-12
ZMO0001	0.57	3.97E-12
ZMO1437	-0.64	5.62E-12
ZMO1936	2.83	1.01E-11
ZMO0693	-0.73	3.09E-11
ZMO1682	0.13	8.02E-11
ZMO0740	-1.08	1.55E-10
ZMO1608	0.11	3.34E-10
ZMO1533	-0.14	4.17E-10
ZMO1366	0.01	6.60E-10

ZMO1873	-0.26	6.82E-10
ZMO1940	2.68	9.44E-10
ZMO0542	-0.26	1.18E-09
ZMO1040	-0.41	1.48E-09
ZMO0571	-0.46	1.62E-09
ZMO0976	-0.76	1.65E-09
ZMO1705	0.04	3.31E-09
ZMO0935	-0.42	3.40E-09
sRNA102	2.79	3.61E-09
ZMOt050	-1.77	4.36E-09
sRNA94	-0.05	4.36E-09
ZMOt033	-0.26	5.39E-09
ZMO1457	0.64	6.42E-09
ZMO1610	-0.75	6.42E-09
ZMO1681	-0.50	7.93E-09
ZMO1609	-0.53	9.51E-09
ZMO1659	-0.05	9.51E-09
ZMO1918	0.35	9.51E-09
ZMO0687	-0.82	1.05E-08
ZMO1007	-0.68	1.11E-08
sRNA104	3.20	1.75E-08
ZMO1289	-0.62	2.87E-08
ZMO1521	-0.41	4.62E-08
ZMO0930	0.65	4.66E-08
ZMO1951	2.82	4.66E-08
ZMO0549	0.58	5.04E-08
ZMO0318	-0.64	5.20E-08
ZMO2030	-0.03	5.27E-08
ZMO1000	0.09	5.33E-08
ZMO0694	-0.51	5.64E-08
ZMO1798	0.53	5.97E-08
ZMO0309	-0.77	8.93E-08
ZMO1075	-0.12	9.12E-08
ZMO1992	-0.19	9.12E-08
ZMO1937	3.79	9.80E-08
ZMO0515	0.45	1.04E-07
ZMO1156	0.71	1.08E-07
ZMO1060	-0.12	1.16E-07
ZMO1945	3.45	1.43E-07
ZMO2003	0.53	1.73E-07

ZMO0929	0.53	1.74E-07
ZMO0611	0.68	2.24E-07
ZMO1915	-0.48	2.34E-07
ZMO1212	-0.46	2.52E-07
ZMO1458	-0.78	3.04E-07
ZMO0667	-0.15	3.86E-07
ZMO1522	0.40	3.86E-07
ZMOt049	-1.93	3.89E-07
ZMO1432	-0.35	5.24E-07
ZMO1581	0.07	5.40E-07
ZMO0174	-0.53	7.10E-07
ZMO1155	0.61	8.54E-07
ZMO0279	0.38	8.92E-07
ZMO0754	-0.45	9.45E-07
ZMO1336	-0.27	9.99E-07
ZMO1844	-0.40	1.06E-06
ZMO0530	0.25	1.07E-06
ZMO0524	0.50	1.29E-06
ZMO1236	-0.52	1.43E-06
ZMO1466	-0.33	1.44E-06
ZMO0928	0.09	1.48E-06
ZMO1076	-0.02	1.52E-06
ZMO0215	-0.24	1.61E-06
ZMO1857	-0.36	1.63E-06
ZMO1917	0.29	1.74E-06
ZMO2005	-0.27	2.56E-06
ZMO1944	2.22	2.93E-06
ZMO0205	-0.57	3.08E-06
ZMO1353	-0.56	3.08E-06
sRNA40	-0.85	3.54E-06
ZMO0379	0.32	3.65E-06
ZMO0516	0.20	3.90E-06
ZMO0578	-0.36	4.15E-06
ZMO1704	0.42	4.48E-06
ZMO0038	-0.34	4.57E-06
ZMO0399	0.30	4.57E-06
ZMO1079	0.53	4.57E-06
ZMO1078	0.18	4.73E-06
ZMO0541	-0.42	5.08E-06
ZMO0728	0.24	5.20E-06

ZMO1225	0.14	5.28E-06
ZMOt020	-2.61	5.31E-06
ZMO0824	-0.25	6.13E-06
ZMO1717	0.05	6.43E-06
ZMO1516	0.79	6.80E-06
ZMO0604	-0.41	6.88E-06
sRNA62	4.70	6.95E-06
ZMO0240	-0.53	7.80E-06
ZMO1787	0.74	8.71E-06
ZMO1596	0.09	9.20E-06
sRNA38	-0.95	1.04E-05
sRNA46	-1.80	1.07E-05
sRNA103	2.34	1.67E-05
ZMO0554	0.55	1.76E-05
ZMO0209	0.51	2.12E-05
ZMO1227	-0.12	2.12E-05
ZMO1477	-1.85	2.20E-05
ZMO0726	-0.50	2.27E-05
ZMO0666	-0.28	2.37E-05
ZMO0948	0.07	2.37E-05
ZMO1831	2.16	2.39E-05
ZMO1590	-0.45	2.71E-05
ZMO0753	-0.26	2.88E-05
ZMO0609	0.70	2.93E-05
sRNA31	0.07	2.93E-05
ZMO1243	1.37	2.97E-05
ZMO1927	-0.25	3.17E-05
ZMO1465	-0.39	3.50E-05
ZMO1751	-1.15	3.62E-05
ZMO2004	0.15	3.62E-05
ZMO0603	0.11	3.64E-05
ZMO1301	-1.66	3.64E-05
ZMO1897	0.01	3.64E-05
ZMOt009	-1.58	3.64E-05
ZMOt021	-2.15	3.66E-05
ZMO1334	-0.88	3.71E-05
ZMO1430	0.02	3.88E-05
ZMO0247	-0.12	4.02E-05
ZMO1588	-0.36	4.12E-05
ZMO0823	-0.41	4.81E-05

ZMO1931	3.08	5.74E-05
ZMO0006	0.28	5.84E-05
ZMO2019	-2.34	5.84E-05
ZMO1932	2.50	6.10E-05
ZMO1943	2.94	6.55E-05
ZMO1459	-1.14	6.75E-05
ZMO1930	2.70	6.75E-05
ZMO0532	0.49	7.57E-05
ZMO0610	0.26	7.59E-05
ZMO1260	-0.38	7.89E-05
ZMO1300	-0.94	8.02E-05
ZMO0520	0.00	8.23E-05
ZMO1014	-1.32	9.41E-05
ZMO0778	-0.13	9.49E-05
ZMO0669	-0.85	9.63E-05
ZMO1081	0.11	9.67E-05
ZMO0229	1.26	1.00E-04
ZMO1429	0.18	1.01E-04
ZMO2034	3.66	1.01E-04
ZMO2014	-0.78	1.02E-04
ZMO1728	-1.22	1.02E-04
ZMO0671	0.00	1.05E-04
ZMO0932	0.00	1.20E-04
sRNA60	-1.73	1.20E-04
ZMO0931	1.33	1.25E-04
ZMO0529	0.11	1.26E-04
ZMO0457	-0.79	1.28E-04
ZMO1863	-0.86	1.33E-04
ZMO0915	-0.72	1.37E-04
ZMO0703	1.09	1.40E-04
ZMO0918	-0.30	1.40E-04
ZMO0400	-0.34	1.44E-04
ZMO1998	0.72	1.44E-04
ZMO0668	-0.45	1.50E-04
ZMO1763	0.25	1.50E-04
ZMO1242	1.28	1.51E-04
ZMO0874	-0.26	1.71E-04
ZMO1910	-0.60	1.71E-04
ZMO0517	0.29	1.92E-04
ZMO1299	-0.33	1.97E-04

ZMO1875	0.06	2.01E-04
ZMO1410	-1.05	2.02E-04
ZMO1950	2.13	2.20E-04
ZMO1215	-0.76	2.21E-04
ZMO0299	-0.26	2.47E-04
ZMO0521	0.14	2.48E-04
ZMO0531	0.15	2.52E-04
ZMOt036	-0.47	2.70E-04
ZMO0096	-0.78	2.73E-04
ZMO0686	-2.16	2.75E-04
ZMO1490	-0.26	3.14E-04
ZMO1582	1.41	3.30E-04
ZMO0605	-0.15	3.55E-04
ZMO0128	0.74	3.67E-04
ZMO0056	0.58	3.70E-04
ZMO1468	-0.95	4.09E-04
ZMO1909	-0.85	4.10E-04
ZMOt007	-4.12	4.16E-04
ZMO0925	-1.39	4.29E-04
ZMO1417	0.47	4.29E-04
ZMO0883	0.35	4.73E-04
ZMO1231	0.33	4.80E-04
ZMO0070	-0.04	4.87E-04
ZMO0362	-0.30	5.09E-04
ZMO1885	0.04	5.12E-04
ZMO0540	0.22	5.63E-04
ZMO0523	0.65	5.89E-04
sRNA75	-1.98	5.95E-04
ZMO1226	-0.20	6.03E-04
ZMO1600	0.13	6.11E-04
sRNA48	1.66	6.19E-04
ZMO1816	0.60	6.33E-04
ZMO1077	-0.19	6.60E-04
ZMO1029	-0.31	6.63E-04
ZMO0467	-0.38	6.94E-04
ZMO1061	0.58	7.12E-04
ZMO1016	0.08	7.17E-04
sRNA69	0.89	7.17E-04
sRNA91	0.81	7.17E-04
Zms3	0.34	7.23E-04

ZMO1425	0.36	7.28E-04
Zms23	-1.71	7.58E-04
ZMO1414	0.01	8.15E-04
ZMO1084	0.06	8.44E-04
ZMO1305	-0.14	8.53E-04
ZMO2011	0.03	8.53E-04
ZMOt030	-2.50	8.53E-04
ZMO0431	1.04	8.83E-04
ZMO0844	0.40	8.83E-04
ZMO1304	-0.07	8.83E-04
ZMO1624	1.84	8.83E-04
ZMO0183	0.88	9.24E-04
ZMO1136	0.29	9.42E-04
ZMOt006	-0.56	9.45E-04
ZMO2015	-0.57	9.95E-04
ZMO0867	-0.26	1.00E-03
ZMO1393	0.23	1.05E-03
ZMO0492	-0.71	1.10E-03
ZMO1209	-0.22	1.20E-03
ZMO0405	0.05	1.23E-03
sRNA32	0.18	1.23E-03
ZMO1352	-0.30	1.26E-03
ZMO0727	0.19	1.27E-03
ZMO1055	-0.71	1.28E-03
ZMO1413	-0.50	1.28E-03
ZMO1654	0.20	1.30E-03
ZMO1636	-0.16	1.31E-03
ZMO0290	0.20	1.38E-03
ZMO1207	0.09	1.38E-03
ZMO1147	0.28	1.38E-03
ZMO0363	0.36	1.40E-03
ZMO0184	-0.74	1.58E-03
ZMO0367	-0.43	1.61E-03
ZMO0380	0.28	1.61E-03
ZMO0796	1.20	1.70E-03
ZMOt003	-1.26	1.70E-03
ZMO0495	-0.44	1.73E-03
ZMO1971	1.65	1.74E-03
ZMO0246	0.34	1.77E-03
ZMO0750	0.19	1.82E-03

ZMO1851	-1.60	1.88E-03
ZMO0526	0.28	2.14E-03
ZMOt002	-2.18	2.20E-03
ZMO1525	0.20	2.26E-03
ZMO1141	-0.24	2.30E-03
ZMO0418	-0.74	2.38E-03
ZMO0113	-0.61	2.44E-03
ZMO0152	0.12	2.45E-03
ZMO1145	-0.10	2.49E-03
ZMO0612	0.49	2.58E-03
ZMO1623	-0.32	2.68E-03
ZMO1673	-0.09	2.75E-03
ZMO1837	0.83	2.78E-03
ZMOt023	1.54	2.95E-03
ZMOt004	-2.11	2.98E-03
ZMO0799	-1.04	3.18E-03
ZMO0965	0.16	3.36E-03
ZMO1780	0.07	3.36E-03
ZMO0732	-0.01	3.47E-03
ZMO1776	-0.29	3.54E-03
ZMO0672	-0.19	3.71E-03
ZMO0613	0.38	4.00E-03
ZMO0714	0.54	4.00E-03
ZMO1903	0.41	4.04E-03
ZMO1625	-0.62	4.28E-03
ZMO1467	0.09	4.31E-03
ZMO0664	-0.28	4.34E-03
ZMO0506	-1.27	4.40E-03
ZMO0179	-0.22	4.43E-03
ZMO0119	-1.03	4.46E-03
ZMO2039	3.18	4.46E-03
ZMO1149	0.01	4.48E-03
ZMO0743	-0.25	4.58E-03
ZMO0591	-0.86	4.76E-03
ZMO0680	-0.93	4.76E-03
ZMO1872	-0.32	4.76E-03
ZMO1821	1.57	4.84E-03
ZMO0884	0.36	5.03E-03
ZMO1855	-0.21	5.16E-03
ZMO1854	0.25	5.17E-03

ZMOt028	-1.43	5.20E-03
ZMO0830	-0.16	5.25E-03
ZMO0134	0.62	5.30E-03
ZMO2017	-0.89	5.41E-03
ZMO0364	-0.59	5.49E-03
ZMO1753	0.04	5.55E-03
ZMO1499	-0.81	5.83E-03
ZMOt010	-1.75	5.83E-03
ZMO2032	1.91	5.83E-03
ZMO0674	-0.51	5.94E-03
ZMO1529	0.05	6.01E-03
ZMO1679	0.05	6.05E-03
ZMO0783	0.57	6.08E-03
ZMO0767	0.51	6.15E-03
sRNA24	-0.58	6.15E-03
ZMO1185	-1.53	6.30E-03
ZMO0528	-0.03	6.52E-03
ZMO1879	-0.01	6.62E-03
sRNA81	-1.31	6.64E-03
ZMO1150	-0.35	6.67E-03
ZMO0908	-0.54	6.85E-03
ZMO1214	0.00	7.11E-03
ZMO0626	-0.60	7.37E-03
ZMO0493	-0.54	7.57E-03
sRNA72	-2.19	7.57E-03
ZMO1105	0.21	7.83E-03
ZMO0253	1.49	7.99E-03
ZMO0150	0.92	8.08E-03
ZMOt029	-1.44	8.08E-03
ZMO0519	0.77	8.28E-03
ZMOt022	0.10	8.52E-03
ZMO0211	0.04	9.05E-03
ZMO0360	-0.33	9.05E-03
sRNA47	1.83	9.14E-03
ZMO1668	-0.51	9.21E-03
ZMO1586	-0.42	9.23E-03
ZMO1622	-0.25	9.25E-03
ZMO1861	-0.38	9.25E-03
ZMO0619	-0.66	9.28E-03
ZMO0062	0.49	9.50E-03

ZMO0538	-0.08	9.75E-03
ZMO0553	-0.55	9.75E-03

GeneID	log2(pZms6/wt)	p_{adj}
Zms6	6.27	1.18E-46
ZMO1928	-3.77	1.28E-21
ZMO0989	-4.36	2.38E-20
ZMO0204	-5.34	6.34E-14
ZMO1929	-2.76	6.34E-14
ZMO0542	3.23	8.64E-13
ZMO1373	-2.23	4.83E-12
ZMO1485	-2.56	3.92E-11
ZMO0728	2.75	3.92E-11
ZMO1424	-3.78	5.29E-11
ZMO1844	-1.95	3.59E-10
ZMO1522	-0.10	5.90E-10
ZMO0976	-3.12	1.67E-09
ZMO0516	2.36	1.75E-09
ZMO0001	0.81	2.18E-09
ZMO0541	2.99	3.73E-09
ZMO1366	3.14	4.39E-09
ZMO1705	-4.46	4.43E-09
ZMO0531	2.76	4.43E-09
ZMO1798	2.12	6.46E-09
ZMO1682	-3.29	7.35E-09
ZMO1681	-4.49	7.71E-09
ZMO1721	-2.01	9.15E-09
ZMO1521	2.72	3.17E-08
ZMO1751	2.86	3.17E-08
ZMO0571	-1.95	3.24E-08
ZMO1918	-3.10	4.93E-08
ZMO0605	2.36	8.41E-08
ZMO0725	3.21	1.24E-07
ZMO1609	-2.14	1.30E-07
ZMO1078	3.36	1.53E-07
ZMO1659	-2.38	1.54E-07
ZMO0387	-6.22	1.54E-07
ZMO2011	1.54	3.40E-07
ZMO0375	-2.16	5.17E-07

ZMO1145	2.39	5.54E-07
ZMO0362	-2.13	6.96E-07
ZMO0610	2.79	7.21E-07
ZMO2003	2.13	8.61E-07
ZMO0521	2.96	1.12E-06
ZMO1581	-3.04	1.40E-06
ZMO1289	1.11	1.80E-06
ZMO2025	1.52	2.32E-06
ZMO1466	-2.63	2.66E-06
ZMO0153	2.30	3.24E-06
ZMO0520	2.75	3.39E-06
ZMO1720	2.16	3.55E-06
ZMO1079	3.14	3.97E-06
ZMO1873	-2.67	4.74E-06
ZMO1305	-2.22	5.56E-06
ZMO0530	3.14	7.38E-06
sRNA31	2.22	8.58E-06
ZMO1155	2.04	1.28E-05
ZMO0222	-2.56	1.49E-05
ZMO1437	-2.90	1.49E-05
ZMO0549	2.27	1.56E-05
ZMO1236	-1.64	1.60E-05
ZMO0606	3.06	1.71E-05
ZMO0532	2.88	1.94E-05
ZMO0554	2.34	2.03E-05
ZMO0515	2.21	2.39E-05
ZMO1955	-1.52	2.40E-05
sRNA100	-3.12	2.40E-05
ZMO0604	1.74	2.42E-05
ZMO0726	3.91	2.42E-05
ZMO0611	2.57	2.52E-05
ZMO0374	-1.40	2.59E-05
ZMO0540	3.32	3.64E-05
ZMO1075	2.67	4.83E-05
ZMO0727	1.85	4.83E-05
ZMO1717	1.28	4.97E-05
ZMO0247	-1.76	5.30E-05
ZMO0612	1.95	6.17E-05
ZMO0529	2.12	6.17E-05
ZMO1025	-2.78	6.96E-05

ZMO1060	-1.45	7.08E-05
ZMO1076	2.61	7.30E-05
ZMO0174	-2.59	0.000112565
ZMO1875	-2.35	0.000112565
ZMO0526	1.43	0.000153016
ZMO1227	1.96	0.00015618
ZMO0318	-1.27	0.000157085
sRNA9	2.25	0.000164333
ZMO2030	-1.36	0.000173929
ZMO0609	2.35	0.000192041
ZMO1279	1.56	0.000202954
ZMO0113	-1.96	0.000211009
ZMO1533	-2.84	0.000218841
ZMO0740	-0.30	0.000221593
ZMOt050	3.81	0.000221593
ZMO0631	2.14	0.000244889
ZMO2013	1.01	0.000267377
Zms15	2.92	0.000282937
ZMO1516	2.17	0.000314946
ZMO2015	1.19	0.000315828
ZMO0732	2.73	0.000405701
ZMO0405	-2.04	0.000527431
ZMO1704	-2.93	0.000608524
ZMO1109	1.08	0.000625257
ZMO0661	-2.10	0.000687892
ZMO0209	0.97	0.000688909
ZMO1593	1.49	0.000688909
ZMO0693	-1.23	0.000770598
ZMO0753	-2.37	0.000786751
ZMO1635	-2.29	0.000786751
ZMO0778	-1.94	0.000842968
ZMOt021	4.77	0.000846495
ZMO0528	1.93	0.000994075
ZMO0883	2.94	0.001083086
ZMO0551	2.37	0.001083086
ZMO0935	-1.33	0.001083086
ZMO1754	-2.20	0.001083086
ZMO0722	1.16	0.001086072
ZMO0309	-2.06	0.001109703
ZMO0948	-1.50	0.001147942

ZMO1845	1.83	0.001147942
ZMO0256	-1.36	0.001175274
ZMO1515	2.09	0.001220064
ZMO0613	1.70	0.001282092
sRNA91	1.00	0.001396668
ZMO1857	-1.10	0.001538464
ZMO1237	0.84	0.001592565
ZMO0687	1.03	0.001641564
ZMO1007	0.37	0.00177902
ZMO0290	-2.40	0.001815915
ZMO0637	1.53	0.00185382
ZMO0743	1.44	0.001933513
ZMOt003	4.77	0.001980085
ZMO1334	-0.72	0.002059311
ZMO0918	-1.73	0.002135555
ZMO1425	-1.65	0.002193955
ZMO2004	3.10	0.002228537
ZMO1077	2.78	0.002246924
ZMO1863	-0.99	0.002261237
ZMO1156	1.79	0.002288712
ZMO0533	3.11	0.002704629
ZMO0524	2.94	0.002790119
ZMO0205	-2.95	0.002790119
ZMO1732	-2.48	0.002940762
sRNA69	-6.46	0.003148065
ZMOt028	3.39	0.003465891
ZMO1413	-1.70	0.003806342
ZMO1246	2.33	0.003882899
ZMO0916	-1.15	0.003891786
ZMO1467	0.38	0.004096071
ZMO0517	2.24	0.004407742
sRNA40	0.23	0.004598359
ZMO1225	2.01	0.004720379
ZMO0454	-1.34	0.00485016
ZMO1040	1.81	0.00485016
ZMO0279	2.01	0.004869927
ZMO0192	-1.56	0.004917748
ZMO0754	-2.39	0.005122289
ZMO0197	-1.58	0.005276337
ZMO0522	3.59	0.005347249

ZMO1608	-1.38	0.005419186
ZMO1992	-2.31	0.00551311
ZMO0179	-1.18	0.005809484
ZMO0038	-1.78	0.006198113
ZMO1226	2.21	0.006198113
ZMO1787	-0.86	0.006198113
ZMO0363	-1.93	0.006468952
ZMOt022	0.39	0.006468952
sRNA32	4.12	0.006468952
ZMO0578	-0.13	0.006643583
ZMO1636	-1.66	0.006643583
ZMO1906	1.27	0.006643583
ZMO0523	2.17	0.006687524
ZMO0694	-1.52	0.006894298
ZMO1669	-0.97	0.006918227
ZMO1759	-1.70	0.006979383
ZMOt014	4.08	0.00729736
ZMO0915	-1.26	0.007774585
ZMO1299	0.62	0.008102653
ZMO0608	1.77	0.008166112
ZMO0833	1.74	0.008166112
ZMO0156	1.14	0.008166112
sRNA70	-0.76	0.008166112
ZMO0677	-0.76	0.009272418
ZMO1993	-2.26	0.009356934

Table A.5: Differentially expressed proteins upon Zms4 and Zms6 induction.

Greater than +/- 2 fold changes from LC-MS/MS normalized spectral counts. Biological duplicates of each strain. Analyzed with Scaffold; Protein threshold: 1.0% FDR, Min peptide: 2, Peptide threshold: 1.0% FDR.

Gene ID	Protein ID	Name	log2(pZms4_5h/4h)
ZMO1096	Q5NNJ0_ZYMMO	Ribonuclease R OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rnr PE=3 SV=1	3.70
ZMO1067	Q5NNL9_ZYMMO	Fe-S metabolism associated SufE OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1067 PE=4 SV=2	3.70

ZMO0211	RL21_ZYMMO	50S ribosomal protein L21 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rplU PE=3 SV=2	3.09
ZMO0963	Q5NNX3_ZYMMO	Transcriptional regulator, TetR family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0963 PE=4 SV=2	2.79
ZMO1086	Q5NNK0_ZYMMO	Glucanase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1086 PE=3 SV=1	2.53
ZMO1516	RL35_ZYMMO	50S ribosomal protein L35 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rpmI PE=3 SV=1	2.43
ZMO0131	Q5NR99_ZYMMO	Metallophosphoesterase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0131 PE=4 SV=2	2.30
ZMO1272	Q5NN14_ZYMMO	Succinylglutamic semialdehyde dehydrogenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1272 PE=3 SV=2	2.13
ZMO1678	Q5NLV8_ZYMMO	Prephenate dehydratase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1678 PE=4 SV=2	1.97
ZMO1303	Q5NMY3_ZYMMO	Pyrroline-5-carboxylate reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=proC PE=3 SV=1	1.88
ZMO1506	Q5NMD0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1506 PE=4 SV=1	1.83
ZMO0760	Q5NPH6_ZYMMO	Glyoxalase/bleomycin resistance protein/dioxygenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0760 PE=4 SV=2	1.79
ZMO1544	Q5NM92_ZYMMO	Cobalt chelatase, pCobS small subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1544 PE=4 SV=1	1.75
ZMO0716	Q5NPM0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0716 PE=4 SV=1	1.73
ZMO1717	Q5NLR9_ZYMMO	TonB family protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1717 PE=4 SV=1	1.70
ZMO0442	Q5NQD9_ZYMMO	HAD-superfamily hydrolase, subfamily IA, variant 3 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0442 PE=4 SV=2	1.68
ZMO1351	Q5NMT5_ZYMMO	Carboxymethylenebutenolidase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1351 PE=4 SV=1	1.68
ZMO0767	Q5NPG9_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0767 PE=4 SV=2	1.64

ZMO1320	Q5NMW6_ZYMMO	Fmu (Sun) domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1320 PE=4 SV=2	1.54
ZMO1688	Q5NLU8_ZYMMO	Folate-binding protein YgfZ OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1688 PE=4 SV=1	1.54
ZMO0226	Q5NR04_ZYMMO	Short-chain dehydrogenase/reductase SDR OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0226 PE=3 SV=1	1.53
ZMO0916	Q5NP20_ZYMMO	Heavy metal transport/detoxification protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0916 PE=4 SV=1	1.53
ZMO0246	HSLV_ZYMMO	ATP-dependent protease subunit HslV OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=hslV PE=3 SV=1	1.51
ZMO1400	Q5NMN6_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1400 PE=4 SV=1	1.49
ZMO1861	D2YW28_ZYMMO	2-nitropropane dioxygenase NPD OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1861 PE=4 SV=1	1.49
ZMO0627	H2VFR0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0627 PE=4 SV=1	1.45
ZMO1022	Q5NNR4_ZYMMO	Cysteine desulfurase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1022 PE=4 SV=1	1.44
ZMO0776	Q5NPG0_ZYMMO	Methyltransferase type 12 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0776 PE=4 SV=1	1.41
ZMO1242	Y1242_ZYMMO	Uncharacterized protein ZMO1242 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1242 PE=4 SV=2	1.34
ZMO0073	Q5NRF7_ZYMMO	CBS domain containing protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0073 PE=4 SV=1	1.33
ZMO0570	Q5NQ11_ZYMMO	Ribosomal L11 methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0570 PE=4 SV=2	1.32
ZMO0834	DDL_ZYMMO	D-alanine--D-alanine ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ddl PE=3 SV=1	1.32
ZMO0286	Q5NQU4_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0286 PE=4 SV=2	1.30
ZMO0551	TRUB_ZYMMO	tRNA pseudouridine synthase B OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=truB PE=3 SV=1	1.30

ZMO1334	H2VFS3_ZYMMO	YceI family protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1334 PE=4 SV=1	1.28
ZMO1409	Q5NMM7_ZYMMO	Inositol monophosphatase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1409 PE=4 SV=2	1.24
ZMO1963	Q5NL23_ZYMMO	Citrate synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1963 PE=3 SV=1	1.23
ZMO1167	Q5NNB9_ZYMMO	Peptidase S15 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1167 PE=4 SV=1	1.23
ZMO1063	Q5NNM3_ZYMMO	Phage shock protein A, PspA OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1063 PE=4 SV=1	1.23
ZMO2025	D9PNM7_ZYMMO	Biopolymer transport protein ExbD/TolR OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO2025 PE=3 SV=1	1.22
ZMO0312	H2VFP4_ZYMMO	Amidohydrolase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0312 PE=4 SV=1	1.19
ZMO0563	Q5NQ18_ZYMMO	Chorismate mutase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0563 PE=4 SV=1	1.19
ZMO1436	Q5NMK0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1436 PE=4 SV=1	1.17
ZMO1760	Q5NLM6_ZYMMO	Extracellular ligand-binding receptor OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1760 PE=4 SV=2	1.15
ZMO0557	RIMP_ZYMMO	Ribosome maturation factor RimP OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rimP PE=3 SV=1	1.14
ZMO0856	EX7S_ZYMMO	Exodeoxyribonuclease 7 small subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=xseB PE=3 SV=1	1.12
ZMO1415	Q5NMM1_ZYMMO	Scaffold protein Nfu/NifU OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1415 PE=4 SV=2	1.09
ZMO1573	Q5NM63_ZYMMO	Dyp-type peroxidase family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1573 PE=4 SV=1	1.08
ZMO2017	D2N0W9_ZYMMO	Capsule biosynthesis phosphatase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO2017 PE=4 SV=1	1.08
ZMO0994	Q5NNU2_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0994 PE=4 SV=2	1.08

ZMO0605	Q5NPX6_ZYMMO	Flagellar hook-associated protein FlgK OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0605 PE=4 SV=1	1.07
ZMO0508	Q5NQ73_ZYMMO	GCN5-related N-acetyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0508 PE=4 SV=1	1.07
ZMO0763	Q5NPH3_ZYMMO	Tetratricopeptide domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0763 PE=4 SV=1	1.07
ZMO0726	RL1_ZYMMO	50S ribosomal protein L1 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rplA PE=3 SV=1	1.06
ZMO0862	Q5NP74_ZYMMO	Nucleoid-associated protein ZMO0862 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0862 PE=3 SV=1	1.06
ZMO1622	Q5NM14_ZYMMO	DNA primase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=dnaG PE=3 SV=2	1.05
ZMO1657	Q5NLX9_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1657 PE=4 SV=1	1.05
ZMO1499	HIS2_ZYMMO	Phosphoribosyl-ATP pyrophosphatase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=hisE PE=3 SV=1	1.04
ZMO0806	Q5NPD0_ZYMMO	Leucyl aminopeptidase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0806 PE=4 SV=1	1.04
ZMO1500	HIS6_ZYMMO	Imidazole glycerol phosphate synthase subunit HisF OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=hisF PE=3 SV=1	1.03
ZMO0757	Q5NPH9_ZYMMO	TPR repeat-containing protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0757 PE=4 SV=2	1.02
ZMO1072	DAPF_ZYMMO	Diaminopimelate epimerase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=dapF PE=1 SV=1	1.02
ZMO1909	Q5NL77_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1909 PE=4 SV=1	1.00
ZMO1675	Q5NLW1_ZYMMO	Protein-disulfide isomerase-like protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1675 PE=4 SV=1	1.00
ZMO1989	Q5NKZ7_ZYMMO	Methylated-DNA/protein-cysteine methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1989 PE=4 SV=1	1.00
ZMO1927	Q5NL59_ZYMMO	Protease 4 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1927 PE=3 SV=1	1.00

ZMO0411	Q5NQG9_ZYMMO	DNA topoisomerase 4 subunit B OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=parE PE=3 SV=1	-1.03
ZMO1543	Q5NM93_ZYMMO	Cobalt chelatase, pCobT subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1543 PE=4 SV=1	-1.03
ZMO1127	Q5NNF9_ZYMMO	tRNA-dihydrouridine synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1127 PE=3 SV=1	-1.04
ZMO1705	Q5NLT1_ZYMMO	Thioredoxin domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1705 PE=4 SV=1	-1.04
ZMO0829	MURD_ZYMMO	UDP-N-acetylmuramoylalanine--D-glutamate ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=murD PE=3 SV=1	-1.06
ZMO1203	Q5NN83_ZYMMO	tRNA (cytidine/uridine-2'-O-)-methyltransferase TrmJ OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=trmJ PE=4 SV=1	-1.09
ZMO1933	Q5NL53_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1933 PE=4 SV=1	-1.13
ZMO1642	ANMK_ZYMMO	Anhydro-N-acetylmuramic acid kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=anmK PE=3 SV=2	-1.15
ZMO1198	Q5NN88_ZYMMO	5-aminolevulinate synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1198 PE=3 SV=2	-1.16
ZMO0270	Q5NQW0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0270 PE=4 SV=1	-1.17
ZMO0937	Q5NNZ9_ZYMMO	Aromatic-amino-acid transaminase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0937 PE=4 SV=2	-1.19
ZMO1188	UBIE_ZYMMO	Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ubiE PE=3 SV=2	-1.23
ZMO0439	Q5NQE2_ZYMMO	Pseudouridine synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0439 PE=3 SV=1	-1.24
ZMO1553	D2YW35_ZYMMO	Thiamine-monophosphate kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1553 PE=4 SV=1	-1.25
ZMO1294	Q5NMZ2_ZYMMO	Sugar isomerase (SIS) OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1294 PE=4 SV=2	-1.27
ZMO1087	Q5NNJ9_ZYMMO	Lytic murein transglycosylase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1087 PE=4 SV=1	-1.27

ZMO1054	Q5NNN2_ZYMMO	DNA topoisomerase 4 subunit A OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=parC PE=3 SV=1	-1.28
ZMO1748	Q5NLN8_ZYMMO	Transcriptional regulator, ArsR family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1748 PE=4 SV=1	-1.29
ZMO1907	MUTS_ZYMMO	DNA mismatch repair protein MutS OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=mutS PE=3 SV=2	-1.37
ZMO1588	UVRA_ZYMMO	UvrABC system protein A OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=uvrA PE=3 SV=2	-1.42
ZMO0079	Q5NRF1_ZYMMO	Response regulator receiver protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0079 PE=4 SV=1	-1.44
ZMO?	Q79RX8_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) PE=4 SV=1	-1.44
ZMO1305	Q5NMY1_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1305 PE=4 SV=1	-1.48
ZMO0094	BIOB_ZYMMO	Biotin synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=bioB PE=3 SV=2	-1.54
ZMO0912	Q5NP24_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0912 PE=4 SV=2	-1.58
ZMO0353	Q5NQM7_ZYMMO	4-diphosphocytidyl-2C-methyl-D-erythritol synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0353 PE=3 SV=2	-1.64
ZMO0964	Q5NNX2_ZYMMO	RND efflux system, outer membrane lipoprotein, NodT family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0964 PE=3 SV=1	-1.66
ZMO0354	Q5NQM6_ZYMMO	DNA mismatch repair protein MutL OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=mutL PE=3 SV=2	-1.68
ZMO1576	Q5NM60_ZYMMO	Short-chain dehydrogenase/reductase SDR OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1576 PE=3 SV=1	-1.72
ZMO1287	Q5NMZ9_ZYMMO	Glycosyl transferase group 1 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1287 PE=4 SV=1	-1.73
ZMO1763	Q5NLM3_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1763 PE=4 SV=2	-1.74
ZMO1530	Q5NMA6_ZYMMO	KpsF/GutQ family protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1530 PE=3 SV=1	-1.74

ZMO1344	Q5NMU2_ZYMMO	Aldo/keto reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1344 PE=4 SV=2	-1.74
ZMO0486	Q5NQ95_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0486 PE=4 SV=1	-1.85
ZMO0176	Q5NR54_ZYMMO	Transketolase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0176 PE=4 SV=1	-1.88
ZMO1925	Q5NL61_ZYMMO	Peptidase U32 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1925 PE=4 SV=1	-1.92
ZMO1954	PANC_ZYMMO	Pantothenate synthetase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=panC PE=3 SV=2	-1.95
ZMO0074	YBEY_ZYMMO	Endoribonuclease YbeY OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ybeY PE=3 SV=1	-2.15
ZMO0573	Q5NQ08_ZYMMO	Glutaredoxin, GrxB family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0573 PE=4 SV=1	-2.16
ZMO1170	Q5NNB6_ZYMMO	Antibiotic biosynthesis monooxygenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1170 PE=4 SV=2	-2.16
ZMO0364	DNLJ_ZYMMO	DNA ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ligA PE=3 SV=2	-2.17
ZMO1271	Q5NN15_ZYMMO	Uroporphyrin-III C-methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1271 PE=3 SV=1	-2.18
ZMO0490	Q5NQ91_ZYMMO	Oligopeptidase B OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0490 PE=4 SV=1	-2.26
ZMO0103	Q5NRC7_ZYMMO	Beta-lactamase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0103 PE=4 SV=2	-2.86
ZMO0007	CYSH_ZYMMO	Phosphoadenosine phosphosulfate reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=cysH PE=3 SV=1	-2.89
ZMO0861	D2YW19_ZYMMO	DNA polymerase III, subunits gamma and tau OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0861 PE=4 SV=1	-3.32
ZMO1590	Q5NM46_ZYMMO	ABC transporter related protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1590 PE=3 SV=1	-3.74

Gene ID	Protein ID	Name	log2(pZms6_5h/4h)

ZMO1569	Q5NM67_ZYMMO	Pyruvate formate-lyase-activating enzyme OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1569 PE=3 SV=2	3.31
ZMO0861	D2YW19_ZYMMO	DNA polymerase III, subunits gamma and tau OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0861 PE=4 SV=1	2.97
ZMO0444	Q5NQD7_ZYMMO	Lipolytic protein G-D-S-L family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0444 PE=4 SV=1	2.82
ZMO0452	Q5NQC9_ZYMMO	Pyrrolo-quinoline quinone OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0452 PE=4 SV=2	2.76
ZMO1067	Q5NNL9_ZYMMO	Fe-S metabolism associated SufE OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1067 PE=4 SV=2	2.64
ZMO0682	Q5NPQ4_ZYMMO	CRISPR-associated protein, Csy1 family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0682 PE=4 SV=1	2.47
ZMO1506	Q5NMD0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1506 PE=4 SV=1	2.33
ZMO0364	DNLJ_ZYMMO	DNA ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ligA PE=3 SV=2	2.30
ZMO0791	PYRB_ZYMMO	Aspartate carbamoyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=pyrB PE=3 SV=1	2.30
ZMO0352	Q5NQM8_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0352 PE=4 SV=2	2.29
ZMO1301	Q5NMY5_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1301 PE=4 SV=1	2.26
ZMO0343	Q5NQN7_ZYMMO	Aminotransferase class IV OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0343 PE=4 SV=1	2.10
ZMO0644	Q5NPT9_ZYMMO	Flagellar motor switch protein FliN OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0644 PE=3 SV=1	2.03
ZMO0785	YQGF_ZYMMO	Putative pre-16S rRNA nuclease OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0785 PE=3 SV=2	2.02
ZMO1642	ANMK_ZYMMO	Anhydro-N-acetylmuramic acid kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=anmK PE=3 SV=2	2.02
ZMO0137	Q5NR93_ZYMMO	Sel1 domain protein repeat-containing protein OS=Zymomonas mobilis subsp. mobilis (strain	1.99

		ATCC 31821 / ZM4 / CP4) GN=ZMO0137 PE=4 SV=1	
ZMO0366	GLF_ZYMMO	Glucose facilitated diffusion protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=glf PE=3 SV=2	1.79
ZMO0908	Q5NP28_ZYMMO	Lipopolysaccharide biosynthesis protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0908 PE=4 SV=2	1.76
ZMO1747	Q5NLN9_ZYMMO	Methylenetetrahydrofolate reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1747 PE=3 SV=1	1.74
ZMO0776	Q5NPG0_ZYMMO	Methyltransferase type 12 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0776 PE=4 SV=1	1.73
ZMO1158	Q5NNC8_ZYMMO	Aspartate racemase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1158 PE=4 SV=1	1.69
ZMO1861	D2YW28_ZYMMO	2-nitropropane dioxygenase NPD OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1861 PE=4 SV=1	1.66
ZMO0286	Q5NQU4_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0286 PE=4 SV=2	1.63
ZMO0298	H2VFR3_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0298 PE=4 SV=1	1.62
ZMO1307	Q5NMX9_ZYMMO	Hydro-lyase, Fe-S type, tartrate/fumarate subfamily, alpha subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1307 PE=4 SV=1	1.62
ZMO1756	GNTP_ZYMMO	Gluconate permease OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=gntP PE=3 SV=1	1.60
ZMO1755	TYSY_ZYMMO	Thymidylate synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=thyA PE=3 SV=3	1.58
ZMO0965	Q5NNX1_ZYMMO	Efflux pump membrane protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0965 PE=4 SV=1	1.58
ZMO1763	Q5NLM3_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1763 PE=4 SV=2	1.58
ZMO1423	Q5NML3_ZYMMO	Peptidase M16 domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1423 PE=3 SV=2	1.57
ZMO0420	TYRC_ZYMMO	Protein TyrC OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=tyrC PE=1 SV=2	1.56

ZMO0208	Q5NR22_ZYMMO	GCN5-related N-acetyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0208 PE=4 SV=1	1.55
ZMO0287	Q5NQU3_ZYMMO	Efflux transporter, RND family, MFP subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0287 PE=4 SV=1	1.55
ZMO0430	Q5NQF1_ZYMMO	Pyrimidine 5'-nucleotidase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0430 PE=4 SV=1	1.55
ZMO1892	Q5NL94_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1892 PE=4 SV=2	1.55
ZMO1468	Q5NMG8_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1468 PE=4 SV=1	1.51
ZMO1093	Q5NNJ3_ZYMMO	Hydrolase, TatD family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1093 PE=4 SV=1	1.50
ZMO0300	EX7L_ZYMMO	Exodeoxyribonuclease 7 large subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=xseA PE=3 SV=2	1.49
ZMO1460	Q5NMH6_ZYMMO	3-mercaptopyruvate sulfurtransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1460 PE=4 SV=1	1.49
ZMO0270	Q5NQW0_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0270 PE=4 SV=1	1.44
ZMO1651	Q5NLY5_ZYMMO	PTSINtr with GAF domain, PtsP OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1651 PE=4 SV=1	1.41
ZMO1562	RPPH_ZYMMO	RNA pyrophosphohydrolase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rppH PE=3 SV=1	1.41
ZMO0215	Q5NR15_ZYMMO	5-formyltetrahydrofolate cyclo-ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0215 PE=3 SV=2	1.41
ZMO1132	LIPA_ZYMMO	Lipoyl synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=lipA PE=3 SV=2	1.40
ZMO0823	RSMH_ZYMMO	Ribosomal RNA small subunit methyltransferase H OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rsmH PE=3 SV=2	1.35
ZMO0158	RUVB_ZYMMO	Holliday junction ATP-dependent DNA helicase RuvB OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ruvB PE=3 SV=1	1.33

ZMO0962	Q5NNX4_ZYMMO	N-acetylglucosamine-6-phosphate deacetylase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0962 PE=3 SV=1	1.33
ZMO1772	D2YW36_ZYMMO	NAD(P)H quinone oxidoreductase, PIG3 family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1772 PE=4 SV=1	1.29
ZMO0503	Q5NQ78_ZYMMO	Glycosyl transferase family 2 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0503 PE=4 SV=1	1.29
ZMO1698	Q5NLT8_ZYMMO	GTP cyclohydrolase-2 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1698 PE=3 SV=2	1.27
ZMO0003	Q5NRM7_ZYMMO	Adenylyl-sulfate kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=cysC PE=3 SV=1	1.27
ZMO1099	Q5NNI7_ZYMMO	Double-strand break repair protein AddB OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1099 PE=4 SV=1	1.26
ZMO0867	Q5NP69_ZYMMO	Hopanoid-associated sugar epimerase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0867 PE=4 SV=2	1.25
ZMO1303	Q5NMY3_ZYMMO	Pyrroline-5-carboxylate reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=proC PE=3 SV=1	1.23
ZMO1703	Q5NLT3_ZYMMO	Ubiquinone biosynthesis hydroxylase, UbiH/UbiF/VisC/COQ6 family OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1703 PE=4 SV=1	1.23
ZMO1013	Y1013_ZYMMO	Maf-like protein ZMO1013 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1013 PE=3 SV=1	1.18
ZMO0442	Q5NQD9_ZYMMO	HAD-superfamily hydrolase, subfamily IA, variant 3 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0442 PE=4 SV=2	1.16
ZMO1090	Q5NNJ6_ZYMMO	Thymidylate kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=tmk PE=3 SV=1	1.13
ZMO1140	Q5NNE6_ZYMMO	Acetolactate synthase, small subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1140 PE=4 SV=1	1.10
ZMO0004	Q5NRM6_ZYMMO	Sulfate adenylyltransferase subunit 1 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=cysN PE=3 SV=1	1.10

ZMO0852	QUEA_ZYMMO	S-adenosylmethionine:tRNA ribosyltransferase-isomerase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=queA PE=3 SV=1	1.10
ZMO0486	Q5NQ95_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0486 PE=4 SV=1	1.06
ZMO0423	Q5NQF7_ZYMMO	FeS assembly protein SufB OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0423 PE=4 SV=1	1.05
ZMO1622	Q5NM14_ZYMMO	DNA primase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=dnaG PE=3 SV=2	1.05
ZMO1900	PLSX_ZYMMO	Phosphate acyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=plsX PE=3 SV=1	1.03
ZMO1170	Q5NNB6_ZYMMO	Antibiotic biosynthesis monooxygenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1170 PE=4 SV=2	1.02
ZMO1320	Q5NMW6_ZYMMO	Fmu (Sun) domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1320 PE=4 SV=2	1.02
ZMO1909	Q5NL77_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1909 PE=4 SV=1	1.00
ZMO0575	Q5NQ06_ZYMMO	Methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0575 PE=4 SV=1	1.00
ZMO0432	Q5NQE9_ZYMMO	Arginase/agmatinase/formiminoglutamase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0432 PE=4 SV=2	-1.00
ZMO0115	Q5NRB5_ZYMMO	Aminotransferase class IV OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0115 PE=3 SV=2	-1.01
ZMO1375	RNC_ZYMMO	Ribonuclease 3 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rnc PE=3 SV=3	-1.02
ZMO0179	Q5NR51_ZYMMO	Fructose-bisphosphate aldolase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0179 PE=4 SV=1	-1.03
ZMO0326	QUEF_ZYMMO	NADPH-dependent 7-cyano-7-deazaguanine reductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=queF PE=3 SV=1	-1.05
ZMO1305	Q5NMY1_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1305 PE=4 SV=1	-1.05

ZMO1928	CH10_ZYMMO	10 kDa chaperonin OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=groS PE=3 SV=2	-1.05
ZMO0490	Q5NQ91_ZYMMO	Oligopeptidase B OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0490 PE=4 SV=1	-1.06
ZMO1712	Q5NLS4_ZYMMO	Peptidyl-prolyl cis-trans isomerase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1712 PE=4 SV=1	-1.08
ZMO0787	Q5NPE9_ZYMMO	Glucose-methanol-choline oxidoreductase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0787 PE=4 SV=1	-1.09
ZMO1753	Q5NLN3_ZYMMO	Oxidoreductase FAD/NAD(P)-binding domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1753 PE=4 SV=1	-1.09
ZMO0149	TRMB_ZYMMO	tRNA (guanine-N(7)-)-methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=trmB PE=3 SV=1	-1.09
ZMO0948	CLPP_ZYMMO	ATP-dependent Clp protease proteolytic subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=clpP PE=3 SV=1	-1.09
ZMO0916	Q5NP20_ZYMMO	Heavy metal transport/detoxification protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0916 PE=4 SV=1	-1.09
ZMO0994	Q5NNU2_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0994 PE=4 SV=2	-1.10
ZMO1576	Q5NM60_ZYMMO	Short-chain dehydrogenase/reductase SDR OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1576 PE=3 SV=1	-1.13
ZMO1553	D2YW35_ZYMMO	Thiamine-monophosphate kinase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1553 PE=4 SV=1	-1.14
ZMO1190	Q5NN96_ZYMMO	Phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1190 PE=4 SV=1	-1.15
ZMO0200	Q5NR30_ZYMMO	Anthranilate phosphoribosyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=trpD PE=3 SV=1	-1.16
ZMO1746	Q5NLP0_ZYMMO	Homocysteine S-methyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1746 PE=4 SV=1	-1.21

ZMO1887	Q5NL99_ZYMMO	Isochorismatase hydrolase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1887 PE=4 SV=1	-1.22
ZMO1780	Q5NLK6_ZYMMO	Antibiotic biosynthesis monooxygenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1780 PE=4 SV=1	-1.24
ZMO1071	Q5NNL5_ZYMMO	MiaB-like tRNA modifying enzyme OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1071 PE=4 SV=1	-1.26
ZMO1511	Q5NMC5_ZYMMO	Uncharacterized protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1511 PE=4 SV=1	-1.30
ZMO1054	Q5NNN2_ZYMMO	DNA topoisomerase 4 subunit A OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=parC PE=3 SV=1	-1.31
ZMO1530	Q5NMA6_ZYMMO	KpsF/GutQ family protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1530 PE=3 SV=1	-1.35
ZMO0607	FLGI_ZYMMO	Flagellar P-ring protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=flgI PE=3 SV=4	-1.36
ZMO0402	RLME_ZYMMO	Ribosomal RNA large subunit methyltransferase E OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=rlmE PE=3 SV=1	-1.37
ZMO1545	Q5NM91_ZYMMO	Heat shock protein DnaJ domain protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1545 PE=4 SV=1	-1.41
ZMO0721	SSRP_ZYMMO	SsrA-binding protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=smpB PE=3 SV=1	-1.44
ZMO0075	Q5NRF5_ZYMMO	PhoH family protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0075 PE=4 SV=2	-1.49
ZMO1502	HIS5_ZYMMO	Imidazole glycerol phosphate synthase subunit HisH OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=hisH PE=3 SV=2	-1.51
ZMO1598	DXS2_ZYMMO	1-deoxy-D-xylulose-5-phosphate synthase 2 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=dxs2 PE=3 SV=2	-1.51
ZMO1904	TSAD_ZYMMO	tRNA N6-adenosine threonylcarbamoyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=tsaD PE=3 SV=1	-1.53
ZMO0863	Q5NP73_ZYMMO	Deoxycytidine triphosphate deaminase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=dcd PE=3 SV=2	-1.55

ZMO0006	CYSG_ZYMMO	Siroheme synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=cysG PE=3 SV=1	-1.65
ZMO1808	Q5NLH8_ZYMMO	UPF0125 protein ZMO1808 OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1808 PE=3 SV=1	-1.73
ZMO1581	Q5NM55_ZYMMO	Antibiotic biosynthesis monooxygenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1581 PE=4 SV=1	-1.86
ZMO0576	Q5NQ05_ZYMMO	Pseudouridine synthase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO0576 PE=3 SV=2	-1.95
ZMO1203	Q5NN83_ZYMMO	tRNA (cytidine/uridine-2'-O-)-methyltransferase TrmJ OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=trmJ PE=4 SV=1	-2.12
ZMO1184	Q5NNA2_ZYMMO	Electron-transferring-flavoprotein dehydrogenase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1184 PE=4 SV=2	-2.16
ZMO1086	Q5NNK0_ZYMMO	Glucanase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1086 PE=3 SV=1	-2.18
ZMO1889	Q5NL97_ZYMMO	SAM-dependent methyltransferase-like protein OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=ZMO1889 PE=4 SV=1	-3.05
ZMO0856	EX7S_ZYMMO	Exodeoxyribonuclease 7 small subunit OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) GN=xseB PE=3 SV=1	-3.30

Table A.6: Plasmids and primers for *Z. mobilis* sRNA studies.

Plasmid name	Sequence	Source
pBBR1MCS2-Pgap		pBBR1MCS2-Pgap-FLP [171]
pEmpty	pBBR1MCS2-SalI-Pgap-NheI	this study
pZms4	pEmpty-NheI-Zms4-BamHI	this study
pZms6	pEmpty-NheI-Zms6-BamHI	this study
pZms18	pEmpty-NheI-Zms18-BamHI	this study
pEZ-Empty		pEZ15Asp [90]
pEZ-Zms4	pEZ-Empty-NcoI-Zms4-SalI	this study
pEZ-Zms6	pEZ-Empty-NcoI-Zms6-SalI	this study
pMS2	pEmpty-NheI-2MS2-BamHI	this study
pMS2-Zms4	pEmpty-NheI-2MS2-Zms4-BamHI	this study

pMS2-Zms6	pEmpty-NheI-2MS2-Zms6-BamHI	this study
pEZ-Zms4-6	pEZ-Zms4-SalI-Zms6	this study
pEZ-Zms4-16	pEZ-Zms4-SalI-Zms16	this study
pEZ-Zms6-16	pEZ-Zms6-SalI-Zms16	this study
pEZ-Zms4-6-16	pEZ-Zms4-6-Zms16	this study

Primers for amplification of EMSA candidates from genome

Target gene	Forward primer sequence	Reverse primer sequence
Zms4	TAGTAATACGACTCACTATAGGGCAC GAGCTCAGAAGTTTTCTGC	GATGAAGTAACAGGTGGTGCG
Zms6	TACGTAA TAC GAC TCA CTA TAG GGA GAT TTT TCT TTT TCA ATC AAC TTT AAT GAT GAA AA	CATATAGTCTTTTGATTACTTCCCCCC GAAAG
Zms16	TACGTAATACGACTCACTATAGGGAG AATAATTTTCCGTTGTTACGGGCTTGC AAACC	AGAGAGGAAGAATCAACCATAGGAC GAAG
sRNA94	TAC GTA ATA CGA CTC ACT ATA GGG AGAATGGGGATGGAGCCATCTCTTTT GTTTTTTATC	TCGTAAAAAAAAGATATCAAGACTG TAGAAAAGCCAAGCC
ZMO0001	TAC GTA ATA CGA CTC ACT ATA GGG AGAATAATATAACATTATTTAATAATT TATATCCAA	AAGTATTACTAACAGTATCAGTCGCG GCTTCTACC
ZMO0170	TAC GTA ATA CGA CTC ACT ATA GGG AGA AATGCGGGCTATACTTTCCATAATATT CCCCAT	AGCTGGGAGCCGCGAAAAGGTC
ZMO0293	TAC GTA ATA CGA CTC ACT ATA GGG AGATCTAAAATATACGCCATTTTTACA TATCGCGTA	GAACGGTACCGAGCAAATATAATTG TGAACCAATAAAATCTG
ZMO0347	TAC GTA ATA CGA CTC ACT ATA GGG AGATAATTGAATAATATTTCATTGTCC TTTGTAATA	GTAATTTGACACCTTTTACCAAAAAC ATCGTCACCGG
ZMO0673	TAC GTA ATA CGA CTC ACT ATA GGG AGATAATTGGAGAAAATTTTATGAGA GTTTGCGCGG	CTGTCAGAAGACCTTGTTCAGTCAAC AGTCTGA
ZMO0693	TAC GTA ATA CGA CTC ACT ATA GGG AGACGATTTCCAGCATCACCTCATTTG CATTATTTC	CCGTCATTTTGATATCTAAGGAACCG TCCGTCG
ZMO0736	TAC GTA ATA CGA CTC ACT ATA GGG AGAAGAAACCGTATAACTATCCGGTT ATTACAGTTT	GTACTTCAATTTCAGTCAGGTTCGTC GCATCC
ZMO1060	TAC GTA ATA CGA CTC ACT ATA GGG AGACATCACAGTGATTGAAAAACAAA ACTGCCCGTT	GCAGGGCGGGTAATGCAAAAGCCAT

ZMO1212	TAC GTA ATA CGA CTC ACT ATA GGG AGAGGCAAAAAAAATATGAATTTCCG ATGAATCTGA	TACTGTCTTCTTCAAAAAGCTGTTTT AAAGTTTTTTCCGAGC
ZMO1437	TAC GTA ATA CGA CTC ACT ATA GGG AGAGTGAAAAAATATACCTCACAATC TCGCTAACAA	GATTGGTGCCAACAGAAAGGACATA TGCCC
ZMO1544	TAC GTA ATA CGA CTC ACT ATA GGG AGATATCACAAGTTATTCTCAAGGAA AGATCATGGC	TATCGGAATCAATGCCGAAGACATC CCTTACC
ZMO1638	TAC GTA ATA CGA CTC ACT ATA GGG AGATCGATAGGTGAAAGCGATATTCA AAAGAGAAGC	AAATTGTTTCCTGACATAATGGCATC CTCGCTG
ZMO1697	TAC GTA ATA CGA CTC ACT ATA GGG AGATTGGGTTTTGGTTTTGGAATATCA ATATCCACC	AAAGGATAACGCCCTTCCATTTGCCA TCA
ZMO1754	TAC GTA ATA CGA CTC ACT ATA GGG AGATTGGAATAGTGGGCTTTGGCTTA AAAAGTAATT	TCGCCGCCCGATCAACGGAATCTTTA AC
ZMO1857	TAC GTA ATA CGA CTC ACT ATA GGG AGATGCTTCACGACGAGAAAGGGGAG CTATCG	AGTTACTAAGGCCTAGTCGTTCAGCT GCTCG
ZMO1863	TAC GTA ATA CGA CTC ACT ATA GGG AGATATATCGCCGTTTTTGCATAAATA AGGACGAAG	TCGCAGCAAAATGATGAGGCGCTGT TTC
Zms7	TAC GTA ATA CGA CTC ACT ATA GGG AGAACACACTGGATGAGTGGGAATCT GGCG	CAATATAGGGATGAGCGCGACCAGC G
ZMO0006	TACGTA ATA CGA CTC ACT ATA GGG AGAGGCCAAAGGCGTTTTACCTAAAC TTTAGGGC	CTTTTTTGAGCAAGGCCACTTTCCGA GCC
ZMO0176	TACGTA ATA CGA CTC ACT ATA GGG AGAATGTTGAGGCTATGCCTCAATAT GG	CGCTATTAGCGGCCTGGATAGC
ZMO0964	TACGTA ATA CGA CTC ACT ATA GGG AGAGGAAGTTTCAGCATATCCTTTTTC AAGGAAAAA	TTACTTTTGGCGCTTTTCCCAATTTGG
ZMO1054	TACGTA ATA CGA CTC ACT ATA GGG AGAGCGCAGTTGACTTTCTGCCTGAT GTCA	AGGTTATGGTTGATAAAGCATAAGC GAGATAGCGTTC
ZMO1312	TACGTA ATA CGA CTC ACT ATA GGG AGAGCCTATCTTGAGATTTTGCTTAAA ATTTATAGT	TATGGACATTCGGGTTGCTTTTAC
ZMO1705	TACGTA ATA CGA CTC ACT ATA GGG AGAGGAAGGGTTATCGTTCCTGTTTCT TTTCTACTA	AATACAGCAAGACCAATTTTGTTTCA GAGGG
ZMO0740	TACGTA ATA CGA CTC ACT ATA GGG AGATAGCAAAGAATCATAAATTTGTT AAATTTTCCT	TCTGCAATTCAGGATCGTCAAT

ZMO0662	TACGTA ATA CGA CTC ACT ATA GGG AGATCAAAAGCGCCAAACTCTCGTCA GAAATAGTCG	ATGCATCGGTAGCATGTGCTTCGATT TCG
ZMO0262	TACGTA ATA CGA CTC ACT ATA GGG AGATCCGGCTTGTTTTGTTAATTCGTC GGGATTGC	AATCCCGCGCACCGCATAATGATCG
ZMO0149	TACGTA ATA CGA CTC ACT ATA GGG AGACAATTCCTCCCGAAAAGAAGGCT TTTCAAGAGG	TTTCCACTAATTCTGCCTGTTTCGG
ZMO0756	TACGTA ATA CGA CTC ACT ATA GGG AGACCCAATCGATAAAAGGCGCTTAT CCTTGCG	CAAATTGGCTTTTCGCCTGTTCGATT ATGCTA
ZMO1934	TACGTA ATA CGA CTC ACT ATA GGG AGAACTTATGCGGACGACAGTGTGTA CACGAC	ACTTATGCGGACGACAGTGTGTACAC GAC
ZMO0208	TACGTA ATA CGA CTC ACT ATA GGG AGATCAGCGTCGATGCCTTGCCATTG G	GGGTTACATCGTTATCGGGAGAGTCG
ZMO1993	TACGTA ATA CGA CTC ACT ATA GGG AGAATAAAGGCAGATCATCCGCCTCT CTTGC	GACCAATAGCCTGATGCCGTAATAA AACTTGG

Other Primers

Primer name	Sequence
2MS2 gblock	GCTAGCCCTGAGGTAATTATAACCCGGGCCCTATATATGGATCCTAAGGT ACCTAATTGCCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTC CAGAAAACATGAGGATCACCCAATGTCTGCAGTATTCCCGGGTTCATTAG ATCTGCGCGCGATCTCTAGA
2MS2_fwd	A TGCCGCTAGCCCTGAGGTAATTATAACC
2MS2_rev	ATGTCTGATATCGATCGCGCGCAGATCTAATGAACCCGGGAA
Zms2-Sal_rev	TCATGTCGACAAACGAAATTGTCTTTCTCTGAAATCG
Nhel-Zms2_fwd	CGATGCTAGCACCAAACCGTAATTGGGGTCGG
Zms3-Sal_rev	TCATGTCGACTGACGGCCCTTACG
Nhel-Zms3_fwd	CGATGCTAGCATGCCCATTTAAAACATCATGAATCCATGTC
Zms8-Sall_rev	TCATGTCGACAAATCTACCTTCCTGCCG
Nhel-Zms8_fwd	CGATGCTAGCAAAATGGAGCAAGAGGAAAAAGAACGACTG
Zms9-Sall_rev	TCATGTCGACTTCTGATATTAATTACA ATACTTAAAAAATTATTTTTACGG
Nhe-Zms9_fwd	CGATGCTAGCGTATTGGATGTTTAATAAGCOGAAGCAGTTCAGG
Zms10-SalI_rev	TCATGTCGACAAGGCGATCATCCTCC
Nhe-Zms10_fwd	CGATGCTAGCCTTTCAGGCGGACAAAAAAGCCG
Zms13-SalI_rev	TCATGTCGACATCCATCAGAAAAAGA GCCG
Nhel-Zms13_fwd	CGATGCTAGCAAAGCCAGTTCAGTTTTGATTGATAAGCTAACAG
Zms14-Sall_rev	TCATGTCGACAGCTGCCTGTGCG
Nhe-Zms14_fwd	CGATGCTAGCATATAAGGTCGCTCTTTTCGAAGAGCGGC
Zms16-Sall_rev	TCATGTCGACA GAGAGGAAGAATCAACCATAGG
Nhel-Zms16_fwd	CGATGCTAGCATAATTTTCCGTTGT TACGGCCTTGCA
Zms20-SalI_rev	TCATGTCGACCCCTTCCAAA GTTGTTCG
Nhel-Zms20_fwd	CGATGCTAGCGTCATACCAGCAAGGOCGCTTAT

UpZms4F	CATGGCATTACCAAACCTGAAATTTTGAAAGTGCG
UpZms4SpeR	ACTTGCTGACCTGCCAATTATTGGGCTGTCAGCTTTTTGGCC
SpeZms4F	TTGGCAGGTCAGCAAGTGCCTGCCCCGATG
SpeZms4DownR	CTTGGACAAAATAAATTACTGGAGCACAGGATGACGCCTAACAAT
DownZms4F	AGTAATTTATTTTGTCCAAGGTGGATTTTTAAACGC
DownZms4R	CCAGATTTTTATAAGGACGCTTTACAATCATACCAC
UpZms6F	GGCCTCTGGCCAGATAATGCCGTTATCGTC
UpZms6SpeR	CGTCATCCTGTGCTCCGTTCAGTTGCAGCAACGGGTTC
SpeZms6F	GGAGCACAGGATGACGCCTAACAATTCATTCAA
SpeZms6DownR	GTCAACCCCCTTGCTCGATTGGCAGGTCAGCAA GTGC
DownZms6F	TCGAGCAAGGGGGTTGA CGCTTACCGTC
DownZms6R	TACCGGATA GCCAAAGATCAAAAATCCCTCTTTT
northern_Zms4	CACAGAAAGCAGGGCAAGGAATTCGGA
northern_Zms6	CCCGAAAGAATCATAAAAAGACTTTA GTCTTTTAGACCA ATCC

Table A.7: Transcripts that co-immunoprecipitate with Zms4 and Zms6.

Values calculated by DESeq2 with biological triplicates of each strain. Transcripts with greater than 1.5-fold enrichment are shown.

GeneID	log2(MS2-Zms4/control)
Zms4	3.20
ZMO1864	1.67
ZMO0262	1.64
sRNA54	1.63
ZMO0204	1.51
ZMO1857	1.49
ZMO0205	1.46
ZMO0263	1.39
ZMO1863	1.29
ZMO0803	1.29
ZMO1851	1.27
ZMO1443	1.21
ZMO0214	1.18
ZMO1648	1.18
sRNA22	1.13
ZMO0695	1.13
ZMO1856	1.13
sRNA55	1.10

sRNA40	1.09
ZMO1917	1.09
ZMO0740	1.08
ZMO0750	1.08
ZMO0174	1.07
ZMO1816	1.06
ZMO2030	1.04
ZMO0133	1.04
ZMO0494	1.03
ZMO1753	1.02
ZMO1463	1.00
ZMO1903	0.99
ZMO1062	0.98
ZMO1973	0.97
ZMO1854	0.97
ZMO1308	0.96
ZMO0989	0.96
ZMO1289	0.96
ZMO0318	0.96
ZMO0275	0.94
ZMO0561	0.94
ZMO0286	0.93
ZMO1705	0.93
ZMO0401	0.92
ZMO1696	0.92
ZMO0801	0.92
ZMO0798	0.92
Zms14	0.91
ZMO1255	0.91
ZMO1776	0.91
ZMO1233	0.91
ZMO1162	0.91
ZMO0267	0.91
ZMO1417	0.91
ZMO0867	0.90
ZMO1335	0.90
ZMO0402	0.90
ZMO1212	0.90
ZMO1810	0.89
Zms7	0.89

ZMO1855	0.89
ZMO0085	0.88
sRNA10	0.88
ZMO0979	0.88
ZMO1697	0.88
ZMO0260	0.87
ZMO0976	0.86
ZMO0231	0.86
ZMO1961	0.86
ZMO0319	0.86
ZMO1447	0.85
ZMO0347	0.85
ZMO0756	0.84
ZMO1026	0.83
ZMO1262	0.83
Zms6	0.83
ZMO1814	0.83
ZMO0124	0.82
ZMO0629	0.82
ZMO0301	0.81
sRNA13	0.81
ZMO0674	0.81
ZMO0889	0.81
ZMO1622	0.80
sRNA37	0.80
ZMO2018	0.80
ZMO1370	0.79
ZMO1041	0.79
ZMO1754	0.79
ZMO0519	0.79
ZMO0733	0.78
Zms9	0.78
ZMO0137	0.78
sRNA21	0.78
ZMO1506	0.78
ZMO1920	0.78
ZMO0089	0.78
sRNA70	0.78
ZMO1115	0.77
ZMO0693	0.77

ZMO0266	0.77
ZMO1475	0.77
ZMO1054	0.77
ZMO0397	0.77
sRNA94	0.75
ZMO0770	0.75
ZMO0343	0.75
ZMO1287	0.74
ZMO0381	0.73
ZMO0090	0.72
ZMO0128	0.71
ZMO1993	0.71
ZMO0176	0.71
ZMO0236	0.71
ZMO0208	0.71
ZMO1522	0.71
ZMO1291	0.71
ZMO0111	0.70
ZMO0172	0.70
ZMO1187	0.70
ZMO0127	0.70
ZMO1787	0.70
ZMO2013	0.69
ZMO2001	0.69
ZMO0978	0.69
ZMO1322	0.69
ZMO0964	0.69
ZMO0121	0.69
ZMO1680	0.69
ZMO1425	0.68
ZMO1265	0.68
ZMO0265	0.68
ZMO0006	0.68
ZMO1524	0.67
ZMO1083	0.67
ZMO0307	0.67
ZMO0271	0.67
ZMO0541	0.67
ZMO0748	0.67
ZMO0375	0.67

ZMO0293	0.67
ZMO1027	0.66
ZMO1918	0.66
ZMO0322	0.66
ZMO2016	0.66
ZMO0996	0.65
ZMO0383	0.65
ZMO0394	0.65
ZMO1428	0.65
ZMO0697	0.65
ZMO1029	0.65
ZMO0857	0.65
ZMO0885	0.65
ZMO0001	0.64
ZMO1618	0.64
ZMO1465	0.64
ZMO1296	0.64
ZMO1698	0.64
ZMO0374	0.64
ZMO0743	0.64
ZMO0662	0.64
sRNA65	0.63
ZMO0930	0.63
ZMO1312	0.63
ZMO0180	0.63
ZMO0095	0.63
ZMO1887	0.63
ZMO1219	0.63
ZMO1324	0.62
ZMO1934	0.62
ZMO0948	0.62
ZMO1779	0.62
Zms17	0.62
ZMO0663	0.62
ZMO1399	0.62
ZMO0436	0.62
ZMO0222	0.61
ZMO1077	0.61
ZMO1060	0.61
ZMO1986	0.61

ZMO1143	0.60
ZMO0038	0.60
ZMO0015	0.60
ZMO1205	0.59
ZMO1079	0.59
ZMO0581	0.59
sRNA28	0.59
ZMO1491	0.59

GeneID	log2(MS2-Zms6/control)
Zms6	2.44
ZMO1452	1.15
ZMO0969	1.13
ZMO0801	1.08
ZMO1810	1.05
ZMO0398	1.04
ZMO1976	0.98
ZMO1544	0.94
ZMO0931	0.92
ZMO0375	0.90
ZMO0399	0.89
ZMO0930	0.87
ZMO0634	0.86
ZMO0629	0.85
ZMO1934	0.82
ZMO1261	0.81
ZMO1284	0.81
ZMO0379	0.80
ZMO1437	0.80
ZMO1457	0.79
ZMO0395	0.78
ZMO1868	0.77
ZMO0584	0.77
ZMO0501	0.77
ZMO1197	0.76
ZMO0002	0.75
ZMO0383	0.74
ZMO0736	0.73
ZMO0381	0.73

sRNA37	0.73
ZMO1291	0.72
ZMO0133	0.71
ZMO0857	0.70
ZMO1252	0.70
ZMO0970	0.70
ZMO0374	0.70
ZMO0546	0.70
ZMO0128	0.70
ZMO0388	0.69
ZMO0406	0.69
ZMO2010	0.69
ZMO0394	0.69
ZMO1156	0.68
ZMO0494	0.67
ZMO0436	0.67
ZMO0756	0.66
ZMO0080	0.65
ZMO0678	0.65
ZMO1809	0.65
ZMO1787	0.64
ZMO0260	0.64
ZMO1288	0.64
ZMO0170	0.63
ZMO0397	0.63
ZMO1255	0.62
ZMO0006	0.62
ZMO0770	0.62
ZMO0380	0.62
ZMO0673	0.61
ZMO0061	0.61
ZMO0149	0.60
Zms16	0.60
ZMO1638	0.60
ZMO1042	0.60
ZMO0932	0.59
ZMO0322	0.59
ZMO0978	0.59
ZMO0255	0.59
ZMO1957	0.59

ZMO1811	0.59

Table A.8: Proteins that co-immunoprecipitate with Zms4 and Zms6.

Biological duplicates of each strain. Analyzed with Scaffold; Protein threshold: 1.0% FDR, Min peptide: 2, Peptide threshold: 1.0% FDR. Fisher's Exact test $p < 0.05$ with Hochberg-Benjamini Correction.

Gene ID	Gene Name	Protein Name	GO	log2(MS2-Zms4/control)
ZMO0753		Glutaredoxin 3	cell [GO:0005623]; electron carrier activity [GO:0009055]; protein disulfide oxidoreductase activity [GO:0015035]; cell redox homeostasis [GO:0045454]	3.46
ZMO0693		OsmC family protein	response to oxidative stress [GO:0006979]	3.17
ZMO1721		Glyoxalase/bleomycin resistance protein/dioxygenase	dioxygenase activity [GO:0051213]	3.00
ZMO1358	rpsT	30S ribosomal protein S20	ribosome [GO:0005840]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	2.91
ZMO2004	rpsS	30S ribosomal protein S19	small ribosomal subunit [GO:0015935]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	2.58
ZMO0531	rplF	50S ribosomal protein L6	ribosome [GO:0005840]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	2.58
ZMO1690		Chaperone DnaJ domain protein	protein folding [GO:0006457]	2.46

ZMO0542	rplQ	50S ribosomal protein L17	ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	2.17
ZMO1609		Uncharacterized protein		1.85
ZMO0994		Uncharacterized protein		1.74
ZMO0178	pgk	Phosphoglycerate kinase (EC 2.7.2.3)	cytoplasm [GO:0005737]; ATP binding [GO:0005524]; phosphoglycerate kinase activity [GO:0004618]; glycolytic process [GO:0006096]	1.58
ZMO1076	rpsP	30S ribosomal protein S16	ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	1.56
ZMO2031	rpmF	50S ribosomal protein L32	large ribosomal subunit [GO:0015934]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	1.46
ZMO0294	rpmB	50S ribosomal protein L28	ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]	1.42
ZMO1305		Uncharacterized protein		1.38
ZMO1147		Outer membrane chaperone Skp (OmpH)		1.22
ZMO1876		Uncharacterized protein		1.09
ZMO0407		GcrA cell cycle regulator		1.09
ZMO0740		CsbD family protein		1.08
ZMO0856	xseB	Exodeoxyribonuclease 7 small subunit (EC 3.1.11.6)	cytoplasm [GO:0005737]; exodeoxyribonuclease VII complex [GO:0009318]; exodeoxyribonuclease VII activity [GO:0008855]; DNA catabolic process [GO:0006308]	1.00

Gene ID	Gene Name	Protein Name	GO		log2(MS2-Zms6/control)
ZMO1657	0	Uncharacterized protein		0	1.70

Table A.9: sRNA candidate regions for *Z. mobilis* oxygen stress.

sRNA	start	end	strand	sRNAscore	mRNAscore	phenoscore	rank
negsRNA119	556193	556393	-	1.000	0.595	1.595	1
possRNA122	744976	745176	+	0.749	0.760	1.508	2
possRNA7	49895	50095	+	0.749	0.757	1.507	3
negsRNA286	1546468	1546668	-	0.561	0.928	1.490	4
possRNA274	1633731	1633931	+	0.794	0.682	1.476	5
negsRNA233	1159638	1159838	-	0.710	0.764	1.474	6
negsRNA302	1616170	1616370	-	0.675	0.790	1.465	7
negsRNA261	1423902	1424102	-	0.490	0.943	1.433	8
possRNA308	1791895	1792095	+	0.629	0.781	1.409	9
possRNA305	1776171	1776580	+	0.574	0.835	1.409	10
possRNA76	518114	518314	+	0.609	0.800	1.409	11
possRNA77	527636	527836	+	0.609	0.790	1.399	12
negsRNA301	1616017	1616217	-	0.550	0.847	1.397	13
negsRNA355	1911885	1912090	-	0.532	0.845	1.377	14
negsRNA333	1796580	1796780	-	0.697	0.677	1.375	15
Zms6	454884	455084	-	0.580	0.781	1.361	16
possRNA100	622360	622560	+	0.524	0.827	1.352	17
possRNA103	622736	622936	+	0.525	0.821	1.346	18
possRNA304	1775982	1776182	+	0.712	0.629	1.341	19
possRNA231	1374154	1374354	+	0.425	0.910	1.334	20
negsRNA70	372219	372419	-	0.402	0.930	1.332	21
negsRNA122	573674	573874	-	0.499	0.824	1.323	22
negsRNA176	749946	750146	-	0.616	0.700	1.317	23
possRNA223	1356300	1356500	+	0.596	0.715	1.311	24
possRNA229	1373490	1373690	+	0.419	0.885	1.304	25

negsRNA175	749848	750048	-	0.661	0.641	1.302	26
negsRNA38	232530	232730	-	0.457	0.838	1.295	27
negsRNA44	234876	235076	-	0.534	0.755	1.289	28
negsRNA310	1644315	1644515	-	0.551	0.736	1.287	29
possRNA212	1286106	1286306	+	0.455	0.830	1.286	30
negsRNA78	406062	406262	-	0.458	0.823	1.281	31
possRNA273	1623560	1623760	+	0.503	0.776	1.279	32
negsRNA202	971258	971458	-	0.550	0.729	1.278	33
negsRNA63	317431	317631	-	0.497	0.781	1.278	34
possRNA21	151281	151481	+	0.488	0.784	1.273	35
negsRNA217	1073447	1073647	-	0.334	0.933	1.266	36
negsRNA75	405412	405637	-	0.493	0.772	1.265	37
negsRNA256	1386312	1386512	-	0.369	0.893	1.261	38
negsRNA180	849424	849624	-	0.494	0.762	1.256	39
negsRNA74	405159	405359	-	0.634	0.614	1.248	40
negsRNA43	234184	234384	-	0.401	0.846	1.247	41
negsRNA251	1368170	1368370	-	0.406	0.841	1.247	42
possRNA13	118321	118521	+	0.406	0.841	1.247	43
negsRNA212	1028554	1028754	-	0.581	0.661	1.242	44
negsRNA324	1746207	1746407	-	0.435	0.806	1.241	45
possRNA2	12581	12781	+	0.485	0.756	1.240	46
negsRNA37	232359	232559	-	0.571	0.668	1.239	47
possRNA14	135565	135765	+	0.537	0.692	1.229	48
possRNA114	691696	691896	+	0.389	0.833	1.222	49
possRNA238	1387667	1387867	+	0.390	0.829	1.219	50
negsRNA338	1826992	1827192	-	0.537	0.669	1.206	51
negsRNA209	1007849	1008049	-	0.368	0.837	1.205	52
Zms8	157567	157724	-	0.535	0.670	1.205	53
negsRNA45	235161	235387	-	0.506	0.699	1.205	54
possRNA260	1563960	1564160	+	0.494	0.710	1.205	55
possRNA157	966480	966680	+	0.514	0.688	1.202	56
possRNA12	117379	117579	+	0.438	0.762	1.200	57
negsRNA76	405765	405965	-	0.461	0.737	1.198	58
possRNA245	1461364	1461564	+	0.354	0.842	1.197	59
negsRNA232	1159477	1159677	-	0.490	0.705	1.195	60

possRNA101	622480	622680	+	0.311	0.884	1.195	61
possRNA232	1374321	1374521	+	0.392	0.803	1.195	62
negsRNA330	1794539	1794739	-	0.258	0.936	1.194	63
negsRNA322	1730188	1730388	-	0.294	0.900	1.193	64
negsRNA94	507977	508177	-	0.427	0.766	1.193	65
negsRNA151	662491	662691	-	0.415	0.778	1.193	66
negsRNA203	971407	971607	-	0.438	0.755	1.192	67
negsRNA148	659937	660137	-	0.476	0.716	1.192	68
possRNA134	754631	754831	+	0.552	0.639	1.191	69
Zms20	258564	258604	+	0.406	0.785	1.191	70
possRNA233	1374585	1374785	+	0.372	0.819	1.191	71
possRNA184	1108914	1109114	+	0.425	0.766	1.191	72
negsRNA127	623826	624026	-	0.390	0.799	1.189	73
negsRNA186	931842	932042	-	0.553	0.630	1.183	74
possRNA311	1795659	1795859	+	0.667	0.512	1.179	75
possRNA34	201393	201593	+	0.272	0.907	1.179	76
negsRNA351	1897899	1898099	-	0.512	0.665	1.177	77
negsRNA291	1564869	1565069	-	0.399	0.777	1.176	78
negsRNA152	663279	663479	-	0.506	0.669	1.175	79
possRNA8	50502	50702	+	0.506	0.668	1.173	80
possRNA275	1644313	1644513	+	0.438	0.733	1.171	81
negsRNA125	617056	617256	-	0.332	0.837	1.169	82
possRNA197	1198122	1198322	+	0.529	0.632	1.161	83
negsRNA80	406877	407077	-	0.414	0.746	1.159	84
negsRNA228	1129317	1129517	-	0.308	0.850	1.158	85
negsRNA280	1529362	1529562	-	0.472	0.685	1.157	86
possRNA237	1387232	1387432	+	0.446	0.710	1.156	87
possRNA115	721549	721749	+	0.490	0.662	1.152	88
negsRNA42	233946	234146	-	0.518	0.634	1.152	89
negsRNA222	1108929	1109129	-	0.374	0.777	1.151	90
negsRNA40	232793	232993	-	0.447	0.704	1.151	91
negsRNA126	622862	623062	-	0.150	1.000	1.150	92
possRNA276	1649585	1649785	+	0.325	0.824	1.149	93
possRNA120	739784	739984	+	0.534	0.614	1.148	94
possRNA123	745105	745305	+	0.531	0.616	1.147	95

negsRNA71	372475	372675	-	0.384	0.759	1.144	96
negsRNA182	875405	875605	-	0.277	0.866	1.143	97
possRNA52	316085	316285	+	0.324	0.818	1.142	98
negsRNA87	423345	423545	-	0.336	0.806	1.142	99
possRNA189	1159634	1159834	+	0.484	0.658	1.141	100
negsRNA47	235839	236039	-	0.465	0.675	1.140	101
negsRNA39	232625	232825	-	0.471	0.669	1.140	102
possRNA306	1776648	1776848	+	0.397	0.741	1.138	103
possRNA215	1300086	1300286	+	0.262	0.876	1.138	104
negsRNA307	1625185	1625385	-	0.325	0.812	1.136	105
negsRNA129	626258	626458	-	0.221	0.915	1.136	106
negsRNA82	407313	407513	-	0.469	0.667	1.136	107
possRNA251	1503319	1503519	+	0.262	0.874	1.136	108
negsRNA299	1612712	1612912	-	0.634	0.500	1.134	109
possRNA104	622860	623060	+	0.324	0.809	1.133	110
negsRNA77	405898	406098	-	0.423	0.707	1.130	111
negsRNA109	522298	522498	-	0.445	0.684	1.129	112
possRNA108	650998	651198	+	0.484	0.645	1.129	113
possRNA65	432691	432891	+	0.255	0.874	1.129	114
negsRNA41	233483	233683	-	0.417	0.708	1.125	115
negsRNA32	215346	215546	-	0.503	0.619	1.121	116
possRNA42	241529	241729	+	0.202	0.917	1.120	117
negsRNA68	370128	370328	-	0.425	0.694	1.120	118
possRNA292	1742462	1742662	+	0.397	0.723	1.119	119
possRNA148	932589	932789	+	0.293	0.824	1.117	120
negsRNA73	374000	374200	-	0.406	0.710	1.116	121
negsRNA196	966231	966431	-	0.195	0.922	1.116	122
negsRNA174	749579	749861	-	0.459	0.656	1.115	123
possRNA57	380353	380553	+	0.425	0.690	1.114	124
negsRNA262	1427452	1427652	-	0.347	0.767	1.114	125
negsRNA2	17599	17799	-	0.369	0.745	1.114	126
possRNA182	1073792	1074238	+	0.211	0.902	1.113	127
possRNA111	656727	656927	+	0.460	0.653	1.113	128
negsRNA35	232160	232360	-	0.411	0.700	1.111	129
negsRNA58	285678	285878	-	0.374	0.732	1.107	130

negsRNA36	232262	232462	-	0.460	0.646	1.106	131
possRNA22	151478	151678	+	0.291	0.814	1.105	132
Zms18	1901190	1901218	-	0.351	0.751	1.101	133
possRNA243	1423947	1424147	+	0.241	0.852	1.094	134
negsRNA239	1233917	1234117	-	0.324	0.767	1.092	135
possRNA17	149886	150086	+	0.311	0.778	1.089	136
possRNA96	611743	611943	+	0.371	0.717	1.088	137
negsRNA213	1046455	1046655	-	0.371	0.716	1.087	138
negsRNA210	1012320	1012520	-	0.159	0.928	1.087	139
negsRNA177	750062	750262	-	0.389	0.696	1.084	140
possRNA319	1810071	1810271	+	0.419	0.663	1.082	141
negsRNA26	182542	182742	-	0.233	0.848	1.081	142
possRNA202	1233918	1234118	+	0.261	0.817	1.079	143
possRNA70	444071	444271	+	0.215	0.859	1.074	144
negsRNA117	540346	540546	-	0.271	0.802	1.073	145
negsRNA282	1540719	1540919	-	0.322	0.749	1.071	146
negsRNA69	371467	371667	-	0.374	0.697	1.071	147
possRNA132	749711	749911	+	0.438	0.631	1.069	148
possRNA277	1649865	1650065	+	0.395	0.673	1.068	149
negsRNA16	149991	150390	-	0.175	0.893	1.068	150
negsRNA135	650975	651175	-	0.146	0.920	1.066	151
possRNA73	467597	467797	+	0.359	0.706	1.065	152
possRNA295	1749843	1750043	+	0.368	0.697	1.065	153
possRNA147	920775	920975	+	0.257	0.808	1.065	154
negsRNA171	745984	746184	-	0.215	0.849	1.064	155
possRNA69	438262	438462	+	0.183	0.881	1.064	156
possRNA241	1416925	1417125	+	0.288	0.775	1.064	157
possRNA235	1386307	1386507	+	0.262	0.800	1.063	158
negsRNA356	1915168	1915408	-	0.305	0.757	1.062	159
possRNA61	405203	405403	+	0.368	0.693	1.062	160
negsRNA79	406366	406676	-	0.331	0.729	1.060	161
possRNA18	150084	150317	+	0.235	0.823	1.058	162
negsRNA303	1618058	1618258	-	0.085	0.972	1.057	163
negsRNA90	444069	444269	-	0.245	0.812	1.056	164
possRNA301	1767373	1767573	+	0.295	0.759	1.054	165

negsRNA100	516085	516285	-	0.176	0.877	1.053	166
possRNA302	1767558	1767758	+	0.310	0.742	1.052	167
possRNA39	214507	214707	+	0.236	0.816	1.051	168
negsRNA240	1241976	1242176	-	0.251	0.799	1.050	169
negsRNA238	1230966	1231166	-	0.406	0.644	1.049	170
possRNA171	1010914	1011114	+	0.235	0.814	1.049	171
possRNA165	1007251	1007451	+	0.256	0.791	1.047	172
negsRNA241	1242258	1242458	-	0.277	0.770	1.047	173
possRNA125	745885	746085	+	0.336	0.710	1.047	174
possRNA143	912484	912684	+	0.355	0.691	1.046	175
possRNA163	1006630	1006830	+	0.404	0.642	1.046	176
possRNA67	437959	438159	+	0.260	0.784	1.043	177
possRNA166	1007454	1007724	+	0.269	0.774	1.043	178
negsRNA249	1356149	1356349	-	0.232	0.810	1.042	179
negsRNA3	17746	17946	-	0.360	0.681	1.041	180
possRNA30	171225	171425	+	0.232	0.805	1.038	181
negsRNA183	888377	888577	-	0.311	0.723	1.035	182
negsRNA30	214816	215016	-	0.249	0.784	1.033	183
negsRNA219	1079738	1079938	-	0.262	0.770	1.033	184
negsRNA67	369742	369942	-	0.421	0.612	1.033	185
negsRNA173	749382	749582	-	0.327	0.703	1.030	186
negsRNA296	1607979	1608179	-	0.448	0.579	1.027	187
possRNA183	1079312	1079512	+	0.334	0.694	1.027	188
negsRNA297	1608290	1608490	-	0.225	0.800	1.025	189
possRNA68	438148	438348	+	0.348	0.676	1.024	190
negsRNA311	1651593	1651793	-	0.195	0.828	1.023	191
possRNA131	749378	749578	+	0.368	0.654	1.022	192
negsRNA121	572567	572767	-	0.444	0.575	1.019	193
possRNA225	1365777	1365977	+	0.527	0.491	1.018	194
possRNA230	1373752	1373952	+	0.266	0.750	1.017	195
possRNA59	382348	382548	+	0.276	0.740	1.016	196
negsRNA66	364110	364310	-	0.021	0.995	1.015	197
negsRNA72	373375	373575	-	0.418	0.591	1.010	198
possRNA282	1659565	1659765	+	0.226	0.782	1.008	199
negsRNA133	638635	638835	-	0.104	0.904	1.008	200

possRNA19	150718	150918	+	0.384	0.623	1.007	201
possRNA91	570829	571029	+	0.157	0.850	1.007	202
negsRNA33	226022	226222	-	0.165	0.841	1.006	203
negsRNA187	932595	932795	-	0.182	0.824	1.006	204
negsRNA332	1796353	1796553	-	0.421	0.584	1.005	205
negsRNA335	1817694	1817894	-	0.087	0.918	1.005	206
negsRNA263	1438402	1438602	-	0.358	0.646	1.004	207
negsRNA46	235539	235739	-	0.439	0.565	1.004	208
possRNA98	622105	622305	+	0.349	0.653	1.002	209
possRNA256	1530017	1530217	+	0.341	0.659	1.000	210
possRNA64	432590	432790	+	0.156	0.842	0.999	211
possRNA167	1007770	1007970	+	0.171	0.827	0.998	212
possRNA208	1258065	1258265	+	0.217	0.779	0.996	213
possRNA152	960030	960230	+	0.173	0.823	0.996	214
negsRNA230	1148762	1148962	-	0.194	0.802	0.996	215
possRNA51	311125	311325	+	0.288	0.707	0.995	216
possRNA278	1650023	1650223	+	0.400	0.594	0.995	217
possRNA28	165909	166109	+	0.209	0.785	0.994	218
possRNA9	76029	76229	+	0.274	0.718	0.992	219
possRNA105	640946	641146	+	0.254	0.738	0.992	220
negsRNA31	215045	215245	-	0.319	0.672	0.991	221
possRNA198	1205569	1205769	+	0.244	0.745	0.988	222
possRNA213	1286279	1286479	+	0.149	0.839	0.988	223
negsRNA64	341953	342153	-	0.179	0.806	0.985	224
possRNA239	1410405	1410605	+	0.226	0.759	0.985	225
possRNA31	174749	174949	+	0.257	0.727	0.983	226
negsRNA260	1423777	1423977	-	0.188	0.794	0.982	227
possRNA53	317087	317287	+	0.231	0.751	0.981	228
negsRNA113	531976	532176	-	0.171	0.810	0.981	229
possRNA204	1241972	1242172	+	0.221	0.759	0.980	230
negsRNA268	1440841	1441041	-	0.212	0.766	0.979	231
negsRNA265	1439565	1439765	-	0.202	0.775	0.977	232
negsRNA292	1578029	1578229	-	0.178	0.797	0.975	233
possRNA62	408416	408616	+	0.275	0.701	0.975	234
negsRNA89	438585	438785	-	0.151	0.824	0.975	235

negsRNA328	1767241	1767441	-	0.304	0.671	0.975	236
negsRNA234	1173059	1173259	-	0.368	0.607	0.975	237
negsRNA227	1128237	1128437	-	0.140	0.835	0.975	238
negsRNA88	438266	438466	-	0.170	0.802	0.973	239
negsRNA146	653012	653212	-	0.110	0.862	0.972	240
negsRNA18	151486	151686	-	0.190	0.781	0.971	241
negsRNA192	961831	962031	-	0.098	0.873	0.970	242
negsRNA231	1159371	1159571	-	0.263	0.708	0.970	243
possRNA224	1365657	1365857	+	0.296	0.673	0.969	244
possRNA35	201562	201762	+	0.144	0.823	0.967	245
negsRNA7	113463	113663	-	0.160	0.807	0.967	246
negsRNA189	943125	943325	-	0.326	0.636	0.962	247
possRNA54	317289	317489	+	0.254	0.708	0.961	248
negsRNA154	690312	690512	-	0.161	0.798	0.959	249
negsRNA270	1461365	1461565	-	0.097	0.862	0.959	250
possRNA139	883540	883740	+	0.199	0.760	0.958	251
possRNA44	282497	282697	+	0.342	0.615	0.957	252
negsRNA272	1493434	1493634	-	0.113	0.842	0.955	253
negsRNA225	1127382	1127582	-	0.124	0.830	0.954	254
negsRNA304	1623568	1623768	-	0.111	0.841	0.952	255
negsRNA327	1766006	1766206	-	0.391	0.560	0.951	256
negsRNA147	656726	656926	-	0.270	0.680	0.950	257
negsRNA65	358104	358304	-	0.149	0.800	0.949	258
negsRNA150	661488	661688	-	0.335	0.613	0.947	259
negsRNA312	1659585	1659785	-	0.171	0.775	0.946	260
possRNA240	1411117	1411317	+	0.199	0.747	0.945	261
possRNA97	621989	622189	+	0.181	0.764	0.945	262
possRNA15	137905	138105	+	0.266	0.679	0.945	263
negsRNA98	514377	514577	-	0.248	0.697	0.944	264
negsRNA143	652464	652664	-	0.123	0.821	0.944	265
negsRNA81	407087	407287	-	0.286	0.657	0.943	266
possRNA41	241064	241264	+	0.098	0.844	0.942	267
negsRNA140	651914	652114	-	0.055	0.886	0.941	268
possRNA86	546023	546223	+	0.160	0.781	0.941	269
possRNA250	1501165	1501365	+	0.176	0.765	0.941	270

possRNA124	745360	745560	+	0.384	0.557	0.941	271
possRNA20	151024	151270	+	0.397	0.543	0.940	272
possRNA222	1356151	1356351	+	0.160	0.779	0.940	273
possRNA129	746572	746772	+	0.150	0.790	0.939	274
negsRNA279	1519893	1520093	-	0.173	0.761	0.934	275
possRNA102	622581	622781	+	0.444	0.488	0.932	276
negsRNA275	1515392	1515592	-	0.041	0.889	0.931	277
possRNA218	1308425	1308625	+	0.113	0.816	0.929	278
negsRNA350	1882364	1882564	-	0.178	0.745	0.923	279
negsRNA271	1491280	1491480	-	0.185	0.736	0.921	280
negsRNA144	652608	652808	-	0.092	0.829	0.921	281
possRNA82	543664	543864	+	0.187	0.734	0.921	282
negsRNA218	1079630	1079830	-	0.115	0.805	0.919	283
negsRNA211	1012662	1012862	-	0.048	0.869	0.917	284
negsRNA178	770684	770884	-	0.143	0.773	0.916	285
possRNA138	875402	875602	+	0.049	0.865	0.913	286
negsRNA185	924475	924675	-	0.186	0.727	0.913	287
possRNA33	201200	201400	+	0.211	0.701	0.913	288
negsRNA344	1841090	1841290	-	0.148	0.765	0.913	289
negsRNA139	651745	651945	-	0.103	0.810	0.913	290
negsRNA341	1840209	1840409	-	0.094	0.817	0.911	291
possRNA45	290477	290677	+	0.339	0.571	0.910	292
negsRNA285	1546059	1546259	-	0.231	0.679	0.910	293
possRNA263	1565140	1565340	+	0.100	0.810	0.910	294
negsRNA145	652834	653034	-	0.229	0.679	0.908	295
negsRNA120	570640	570840	-	0.239	0.668	0.906	296
possRNA90	570643	570843	+	0.158	0.748	0.906	297
negsRNA337	1821273	1821473	-	0.002	0.904	0.906	298
negsRNA60	297696	297896	-	0.178	0.727	0.905	299
possRNA279	1654267	1654467	+	0.203	0.701	0.904	300
possRNA126	745980	746180	+	0.087	0.817	0.904	301
possRNA135	770684	770884	+	0.173	0.731	0.903	302
negsRNA293	1580873	1581073	-	0.154	0.749	0.903	303
possRNA246	1465845	1466045	+	0.028	0.875	0.903	304
possRNA158	971248	971448	+	0.137	0.765	0.901	305

possRNA37	202039	202239	+	0.145	0.754	0.900	306
negsRNA103	519420	519620	-	0.195	0.704	0.899	307
negsRNA110	524397	524597	-	0.186	0.713	0.899	308
possRNA205	1242256	1242456	+	0.035	0.864	0.899	309
negsRNA118	554108	554308	-	0.231	0.667	0.898	310
negsRNA172	747033	747233	-	0.255	0.643	0.898	311
negsRNA107	521169	521369	-	0.106	0.791	0.897	312
negsRNA96	513093	513293	-	0.062	0.835	0.896	313
possRNA264	1565491	1565691	+	0.345	0.551	0.896	314
negsRNA316	1706364	1706564	-	0.250	0.646	0.896	315
negsRNA300	1612965	1613165	-	0.245	0.649	0.894	316
possRNA71	454879	455079	+	0.126	0.768	0.894	317
possRNA25	157546	157746	+	0.196	0.698	0.894	318
negsRNA214	1049909	1050109	-	0.184	0.710	0.894	319
possRNA315	1797424	1797624	+	0.027	0.866	0.892	320
negsRNA252	1381201	1381401	-	0.039	0.852	0.891	321
possRNA161	992290	992490	+	0.123	0.768	0.891	322
possRNA145	913029	913229	+	0.214	0.676	0.890	323
possRNA144	912669	912869	+	0.085	0.805	0.890	324
negsRNA21	171254	171454	-	0.090	0.800	0.890	325
possRNA300	1767257	1767457	+	0.217	0.672	0.889	326
negsRNA290	1564694	1564894	-	0.217	0.672	0.888	327
possRNA247	1478386	1478586	+	0.040	0.848	0.888	328
negsRNA216	1062394	1062594	-	0.157	0.731	0.888	329
possRNA121	743518	743718	+	0.073	0.814	0.888	330
negsRNA325	1748898	1749098	-	0.092	0.795	0.887	331
negsRNA181	868905	869105	-	0.191	0.695	0.886	332
negsRNA184	920781	920981	-	0.042	0.841	0.884	333
negsRNA48	239216	239416	-	0.226	0.657	0.883	334
negsRNA156	721548	721748	-	0.096	0.786	0.882	335
negsRNA57	283268	283468	-	0.078	0.803	0.882	336
negsRNA195	966027	966227	-	0.181	0.701	0.881	337
negsRNA101	516784	516984	-	0.198	0.683	0.880	338
possRNA32	182540	182740	+	0.033	0.847	0.880	339
negsRNA323	1737440	1737640	-	0.068	0.811	0.880	340

negsRNA6	96315	96515	-	0.047	0.831	0.878	341
negsRNA317	1706552	1706752	-	0.214	0.663	0.876	342
negsRNA11	139674	139874	-	0.215	0.661	0.876	343
possRNA130	746984	747235	+	0.091	0.783	0.873	344
possRNA320	1810600	1810800	+	0.151	0.722	0.873	345
Zms3	513151	513223	+	0.167	0.705	0.873	346
negsRNA208	1007253	1007453	-	0.115	0.758	0.873	347
negsRNA244	1294764	1294964	-	0.064	0.808	0.872	348
possRNA55	317430	317630	+	0.102	0.770	0.872	349
negsRNA336	1818796	1818996	-	0.135	0.737	0.872	350
negsRNA29	214508	214708	-	0.060	0.812	0.872	351
possRNA26	157708	157908	+	0.098	0.773	0.871	352
negsRNA346	1841407	1841607	-	0.098	0.773	0.871	353
negsRNA17	150656	150856	-	0.153	0.718	0.871	354
negsRNA104	520486	520686	-	0.197	0.673	0.870	355
possRNA149	942931	943131	+	0.153	0.716	0.869	356
negsRNA205	1006356	1006556	-	0.088	0.779	0.867	357
negsRNA179	841062	841262	-	0.066	0.798	0.865	358
Zms24	1607592	1607718	+	0.226	0.639	0.864	359
possRNA296	1755843	1756043	+	0.034	0.830	0.864	360
possRNA48	290919	291119	+	0.173	0.688	0.862	361
negsRNA284	1541419	1541619	-	0.197	0.664	0.862	362
negsRNA278	1515916	1516116	-	0.119	0.742	0.862	363
negsRNA22	171426	171626	-	0.127	0.734	0.862	364
possRNA179	1056277	1056477	+	0.124	0.735	0.859	365
possRNA16	138778	138978	+	0.115	0.744	0.858	366
negsRNA226	1127523	1127723	-	0.133	0.726	0.858	367
negsRNA52	242307	242507	-	0.222	0.636	0.858	368
negsRNA53	258486	258686	-	0.066	0.792	0.858	369
negsRNA267	1440413	1440613	-	0.039	0.818	0.857	370
possRNA187	1138695	1138895	+	0.232	0.624	0.856	371
possRNA322	1817696	1817896	+	0.044	0.812	0.856	372
negsRNA159	725581	725781	-	0.124	0.731	0.855	373
negsRNA246	1345927	1346127	-	0.085	0.769	0.854	374
possRNA211	1262556	1262756	+	0.132	0.722	0.854	375

negsRNA84	409918	410118	-	0.065	0.790	0.854	376
possRNA11	113492	113692	+	0.219	0.634	0.853	377
negsRNA318	1707214	1707414	-	0.128	0.723	0.852	378
negsRNA247	1350494	1350694	-	0.073	0.778	0.851	379
possRNA60	382545	382745	+	0.210	0.641	0.851	380
negsRNA223	1127003	1127203	-	0.010	0.840	0.850	381
possRNA156	965723	965923	+	0.181	0.669	0.849	382
negsRNA194	965727	965927	-	0.026	0.822	0.848	383
negsRNA61	311129	311329	-	0.193	0.655	0.848	384
negsRNA245	1334927	1335127	-	0.173	0.675	0.848	385
possRNA74	475555	475755	+	0.036	0.812	0.847	386
negsRNA334	1801377	1801577	-	0.069	0.778	0.847	387
possRNA85	545843	546043	+	0.121	0.724	0.846	388
possRNA175	1047722	1047922	+	0.049	0.797	0.846	389
possRNA234	1384168	1384368	+	0.089	0.757	0.845	390
possRNA168	1010223	1010423	+	0.108	0.736	0.844	391
possRNA253	1513870	1514070	+	0.125	0.718	0.843	392
negsRNA112	527634	527834	-	0.012	0.831	0.843	393
possRNA214	1294551	1294751	+	0.077	0.765	0.842	394
negsRNA347	1841788	1841988	-	0.112	0.727	0.840	395
negsRNA206	1006586	1006786	-	0.146	0.693	0.839	396
possRNA203	1235592	1235792	+	0.215	0.624	0.839	397
negsRNA51	242159	242359	-	0.148	0.690	0.839	398
negsRNA235	1205568	1205768	-	0.116	0.722	0.838	399
possRNA281	1659219	1659419	+	0.159	0.679	0.838	400
possRNA257	1540857	1541057	+	0.037	0.798	0.836	401
negsRNA142	652238	652438	-	0.028	0.808	0.836	402
possRNA317	1797882	1798082	+	0.144	0.691	0.835	403
negsRNA102	518107	518307	-	0.069	0.765	0.835	404
possRNA136	818165	818365	+	0.035	0.799	0.834	405
possRNA133	750061	750261	+	0.139	0.695	0.834	406
negsRNA188	942939	943139	-	0.029	0.804	0.833	407
possRNA83	544428	544628	+	0.255	0.578	0.833	408
negsRNA161	726699	726899	-	0.006	0.826	0.832	409
negsRNA169	743510	743710	-	0.034	0.797	0.830	410

possRNA194	1173410	1173610	+	0.111	0.718	0.829	411
possRNA191	1162379	1162579	+	0.181	0.647	0.828	412
negsRNA59	291611	291811	-	0.139	0.688	0.827	413
negsRNA200	967606	967806	-	0.117	0.709	0.826	414
negsRNA221	1086366	1086566	-	0.123	0.702	0.825	415
possRNA79	540348	540548	+	0.061	0.763	0.824	416
possRNA286	1700110	1700310	+	0.089	0.736	0.824	417
possRNA221	1351035	1351235	+	0.097	0.727	0.824	418
negsRNA277	1515751	1515951	-	0.111	0.711	0.822	419
negsRNA345	1841163	1841363	-	0.019	0.803	0.822	420
negsRNA273	1510277	1510477	-	0.176	0.646	0.822	421
negsRNA237	1223858	1224058	-	0.057	0.765	0.822	422
possRNA107	644647	644847	+	0.022	0.799	0.821	423
possRNA47	290786	290986	+	0.078	0.742	0.821	424
negsRNA124	597672	597872	-	0.172	0.648	0.820	425
Zms4	1350887	1351015	-	0.101	0.719	0.820	426
possRNA106	641164	641364	+	0.083	0.736	0.819	427
negsRNA158	725301	725501	-	0.132	0.687	0.819	428
possRNA255	1529839	1530039	+	0.208	0.610	0.817	429
negsRNA5	89901	90101	-	0.019	0.798	0.817	430
negsRNA95	511442	511642	-	0.039	0.778	0.816	431
possRNA36	201821	202021	+	0.082	0.734	0.816	432
possRNA269	1612879	1613079	+	0.164	0.652	0.816	433
possRNA249	1498519	1498719	+	0.172	0.643	0.815	434
negsRNA191	960034	960234	-	0.006	0.808	0.814	435
negsRNA352	1898118	1898318	-	0.203	0.610	0.813	436
negsRNA308	1627428	1627628	-	0.013	0.800	0.813	437
negsRNA320	1729243	1729443	-	0.100	0.713	0.813	438
possRNA270	1615941	1616141	+	0.169	0.643	0.813	439
negsRNA128	625776	625976	-	0.082	0.730	0.812	440
negsRNA329	1776654	1776854	-	0.050	0.763	0.812	441
possRNA24	157313	157513	+	0.286	0.526	0.812	442
negsRNA199	967327	967527	-	0.081	0.730	0.811	443
negsRNA340	1837528	1837728	-	0.178	0.633	0.811	444
possRNA109	651185	651385	+	0.174	0.636	0.810	445

possRNA84	544564	544764	+	0.115	0.695	0.810	446
possRNA248	1491273	1491473	+	0.136	0.673	0.810	447
negsRNA83	409110	409310	-	0.115	0.695	0.809	448
possRNA210	1262471	1262671	+	0.089	0.720	0.809	449
possRNA299	1762790	1762990	+	0.079	0.730	0.809	450
possRNA58	381204	381404	+	0.098	0.711	0.808	451
negsRNA339	1834268	1834468	-	0.068	0.739	0.808	452
possRNA190	1162034	1162234	+	0.150	0.657	0.807	453
negsRNA224	1127193	1127393	-	0.024	0.783	0.806	454
negsRNA207	1007013	1007213	-	0.135	0.671	0.806	455
possRNA160	991974	992174	+	0.164	0.642	0.806	456
possRNA290	1706254	1706454	+	0.090	0.715	0.806	457
possRNA287	1700963	1701163	+	0.134	0.672	0.806	458
negsRNA197	966483	966683	-	0.009	0.797	0.805	459
negsRNA243	1260334	1260534	-	0.017	0.787	0.804	460
possRNA6	40182	40382	+	0.117	0.687	0.804	461
negsRNA309	1633730	1633930	-	0.051	0.752	0.804	462
possRNA209	1259689	1259889	+	0.213	0.590	0.802	463
Zms16	868984	869017	+	0.142	0.661	0.802	464
negsRNA86	414861	415061	-	0.088	0.714	0.802	465
possRNA316	1797567	1797767	+	0.000	0.801	0.802	466
negsRNA283	1541151	1541351	-	0.035	0.765	0.800	467
possRNA258	1541103	1541303	+	0.011	0.789	0.800	468
possRNA89	559391	559591	+	0.108	0.691	0.799	469
negsRNA114	532647	532847	-	0.009	0.790	0.799	470
possRNA119	737566	737766	+	0.092	0.705	0.797	471
negsRNA342	1840375	1840575	-	0.019	0.777	0.797	472
possRNA150	946685	946885	+	0.038	0.759	0.796	473
possRNA95	610759	610959	+	0.169	0.627	0.796	474
negsRNA315	1706259	1706459	-	0.227	0.569	0.796	475
negsRNA25	174752	174952	-	0.052	0.744	0.795	476
negsRNA91	444502	444702	-	0.078	0.715	0.794	477
possRNA298	1761119	1761319	+	0.024	0.768	0.791	478
negsRNA49	239504	239704	-	0.020	0.770	0.790	479
possRNA266	1598120	1598320	+	0.035	0.755	0.790	480

possRNA186	1129316	1129516	+	0.003	0.786	0.789	481
possRNA337	1915194	1915406	+	0.012	0.777	0.789	482
possRNA314	1797325	1797525	+	0.010	0.778	0.788	483
possRNA244	1457000	1457200	+	0.014	0.774	0.788	484
negsRNA264	1438991	1439191	-	0.029	0.760	0.788	485
negsRNA266	1439946	1440146	-	0.074	0.714	0.788	486
negsRNA132	637598	637798	-	0.019	0.767	0.786	487
possRNA220	1334922	1335122	+	0.142	0.643	0.785	488
negsRNA229	1138695	1138895	-	0.030	0.754	0.784	489
possRNA118	737450	737650	+	0.085	0.697	0.781	490
negsRNA141	652058	652258	-	0.045	0.736	0.781	491
possRNA5	39924	40124	+	0.016	0.765	0.781	492
negsRNA255	1385936	1386136	-	0.020	0.761	0.781	493
negsRNA15	148744	148944	-	0.036	0.744	0.780	494
negsRNA10	138785	138985	-	0.044	0.736	0.779	495
possRNA201	1231931	1232131	+	0.171	0.607	0.778	496
possRNA46	290597	290797	+	0.149	0.630	0.778	497
possRNA188	1148532	1148732	+	0.117	0.660	0.777	498
possRNA185	1126914	1127114	+	0.005	0.773	0.777	499
negsRNA242	1253364	1253564	-	0.041	0.736	0.777	500
negsRNA306	1624021	1624221	-	0.020	0.756	0.777	501
possRNA321	1815298	1815498	+	0.078	0.698	0.776	502
possRNA116	721725	721925	+	0.000	0.776	0.776	503
negsRNA276	1515612	1515812	-	0.090	0.685	0.776	504
possRNA38	214289	214489	+	0.036	0.740	0.776	505
possRNA178	1049910	1050110	+	0.050	0.725	0.775	506
possRNA177	1048052	1048252	+	0.053	0.722	0.775	507
possRNA78	540147	540353	+	0.045	0.730	0.775	508
negsRNA50	241064	241264	-	0.020	0.753	0.773	509
possRNA29	166783	166983	+	0.162	0.612	0.773	510
negsRNA157	722823	723023	-	0.115	0.655	0.770	511
negsRNA134	640760	640960	-	0.078	0.692	0.770	512
negsRNA4	18274	18474	-	0.095	0.674	0.770	513
possRNA113	690133	690333	+	0.136	0.633	0.769	514
possRNA280	1654506	1654706	+	0.096	0.671	0.767	515

possRNA49	297500	297700	+	0.181	0.586	0.767	516
negsRNA123	596826	597026	-	0.110	0.656	0.766	517
possRNA288	1701138	1701338	+	0.010	0.756	0.765	518
Zms10	39468	39504	+	0.125	0.639	0.764	519
negsRNA201	971155	971355	-	0.221	0.543	0.764	520
possRNA272	1623456	1623656	+	0.106	0.656	0.762	521
negsRNA254	1385814	1386014	-	0.075	0.686	0.761	522
possRNA1	11555	11755	+	0.076	0.686	0.761	523
possRNA236	1387131	1387331	+	0.031	0.731	0.761	524
negsRNA160	726462	726662	-	0.042	0.718	0.760	525
possRNA140	884868	885068	+	0.063	0.697	0.760	526
negsRNA166	730212	730412	-	0.034	0.726	0.760	527
possRNA99	622227	622427	+	0.176	0.583	0.759	528
possRNA87	553449	553649	+	0.032	0.727	0.759	529
possRNA66	432989	433189	+	0.017	0.741	0.759	530
possRNA206	1246987	1247187	+	0.012	0.747	0.758	531
negsRNA23	171919	172119	-	0.035	0.723	0.758	532
possRNA242	1417461	1417661	+	0.149	0.609	0.758	533
possRNA291	1737430	1737630	+	0.035	0.722	0.756	534
negsRNA8	125299	125499	-	0.023	0.733	0.756	535
possRNA180	1062394	1062594	+	0.025	0.731	0.756	536
negsRNA162	726830	727030	-	0.011	0.745	0.756	537
negsRNA289	1563964	1564164	-	0.166	0.589	0.756	538
possRNA271	1616161	1616361	+	0.040	0.715	0.754	539
negsRNA131	630457	630657	-	0.109	0.644	0.753	540
possRNA313	1797134	1797334	+	0.025	0.728	0.753	541
possRNA92	588738	588938	+	0.034	0.719	0.753	542
possRNA40	227399	227599	+	0.013	0.738	0.750	543
negsRNA319	1707452	1707652	-	0.041	0.710	0.750	544
possRNA285	1692292	1692492	+	0.040	0.710	0.750	545
possRNA169	1010475	1010675	+	0.145	0.605	0.749	546
negsRNA108	521863	522063	-	0.184	0.564	0.748	547
possRNA164	1007013	1007213	+	0.087	0.660	0.747	548
possRNA217	1303263	1303463	+	0.096	0.651	0.747	549
negsRNA168	739749	739949	-	0.132	0.614	0.746	550

negsRNA153	679162	679362	-	0.100	0.646	0.746	551
possRNA200	1230966	1231166	+	0.110	0.635	0.745	552
negsRNA343	1840872	1841072	-	0.116	0.628	0.744	553
negsRNA204	991980	992180	-	0.054	0.690	0.744	554
negsRNA170	745897	746097	-	0.035	0.708	0.743	555
possRNA81	543552	543752	+	0.105	0.637	0.742	556
negsRNA12	148033	148233	-	0.094	0.646	0.741	557
negsRNA9	125513	125713	-	0.038	0.702	0.740	558
possRNA310	1794539	1794739	+	0.013	0.727	0.740	559
possRNA195	1173712	1173912	+	0.022	0.717	0.739	560
possRNA265	1593617	1593817	+	0.181	0.556	0.738	561
possRNA93	603759	603959	+	0.053	0.684	0.737	562
negsRNA215	1056277	1056477	-	0.000	0.735	0.736	563
possRNA174	1040135	1040335	+	0.016	0.719	0.735	564
negsRNA105	520780	520980	-	0.056	0.677	0.734	565
negsRNA1	12645	12845	-	0.000	0.731	0.731	566
negsRNA295	1607559	1607759	-	0.038	0.691	0.729	567
negsRNA165	729280	729480	-	0.018	0.711	0.729	568
possRNA181	1073528	1073729	+	0.074	0.655	0.729	569
negsRNA288	1563639	1563839	-	0.043	0.685	0.728	570
possRNA297	1760952	1761152	+	0.025	0.703	0.728	571
possRNA94	610043	610243	+	0.031	0.697	0.728	572
negsRNA250	1356301	1356501	-	0.025	0.698	0.724	573
possRNA27	158085	158285	+	0.097	0.626	0.723	574
possRNA88	554100	554300	+	0.041	0.679	0.721	575
negsRNA116	540149	540350	-	0.116	0.604	0.720	576
possRNA227	1367576	1367776	+	0.024	0.694	0.718	577
possRNA289	1702957	1703157	+	0.107	0.611	0.718	578
negsRNA149	660413	660613	-	0.019	0.697	0.717	579
possRNA72	462321	462521	+	0.119	0.597	0.716	580
negsRNA220	1085085	1085285	-	0.065	0.651	0.716	581
possRNA142	890260	890460	+	0.025	0.690	0.715	582
negsRNA28	201665	201865	-	0.029	0.685	0.715	583
possRNA146	913381	913581	+	0.019	0.695	0.714	584
negsRNA298	1612407	1612607	-	0.245	0.469	0.714	585

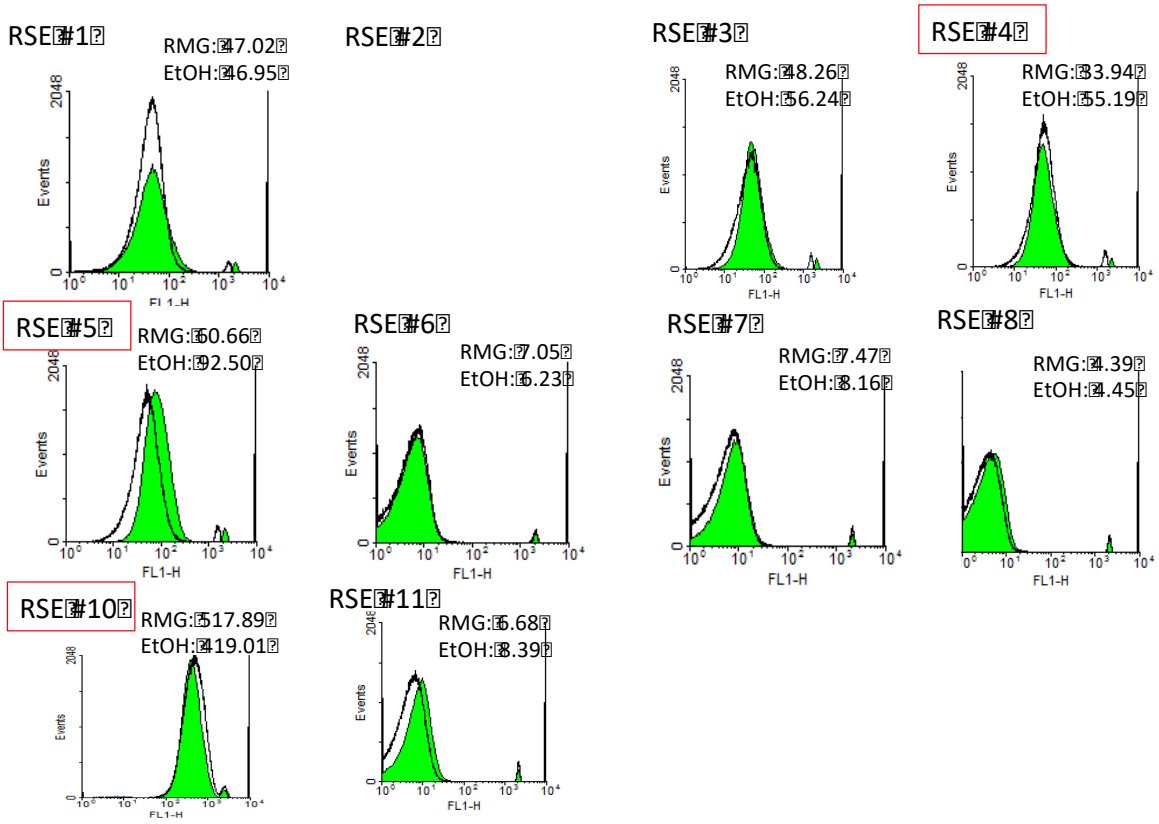
negsRNA34	227404	227604	-	0.067	0.646	0.714	586
possRNA303	1768189	1768389	+	0.087	0.626	0.713	587
negsRNA14	148616	148816	-	0.065	0.647	0.712	588
possRNA309	1793464	1793664	+	0.016	0.696	0.712	589
negsRNA62	317298	317498	-	0.089	0.620	0.709	590
negsRNA27	194830	195030	-	0.087	0.622	0.709	591
possRNA173	1028552	1028752	+	0.058	0.649	0.708	592
negsRNA136	651188	651388	-	0.002	0.705	0.707	593
possRNA170	1010651	1010851	+	0.021	0.685	0.706	594
possRNA193	1173101	1173301	+	0.059	0.647	0.706	595
possRNA294	1748897	1749097	+	0.063	0.642	0.705	596
negsRNA163	727282	727482	-	0.017	0.688	0.705	597
negsRNA314	1667244	1667444	-	0.026	0.679	0.704	598
negsRNA258	1411142	1411342	-	0.009	0.695	0.704	599
possRNA207	1253397	1253597	+	0.097	0.606	0.703	600
possRNA50	297711	297911	+	0.023	0.680	0.703	601
possRNA176	1047856	1048056	+	0.041	0.660	0.701	602
possRNA141	884990	885190	+	0.028	0.673	0.701	603
negsRNA97	514030	514230	-	0.054	0.647	0.701	604
possRNA261	1564701	1564901	+	0.013	0.685	0.698	605
negsRNA93	476679	476879	-	0.018	0.679	0.698	606
possRNA117	725303	725503	+	0.035	0.663	0.698	607
possRNA112	689899	690099	+	0.145	0.552	0.697	608
possRNA4	39647	39847	+	0.027	0.668	0.696	609
negsRNA20	166786	166986	-	0.089	0.606	0.695	610
negsRNA321	1730104	1730304	-	0.026	0.668	0.694	611
possRNA312	1796740	1796940	+	0.132	0.559	0.691	612
negsRNA167	739491	739691	-	0.011	0.680	0.691	613
possRNA10	110440	110640	+	0.075	0.614	0.690	614
possRNA254	1515758	1515958	+	0.003	0.687	0.690	615
Zms15	1666935	1666966	+	0.000	0.688	0.688	616
negsRNA55	258846	259046	-	0.017	0.671	0.687	617
possRNA162	993011	993211	+	0.056	0.631	0.687	618
possRNA155	963018	963218	+	0.069	0.615	0.684	619
negsRNA193	963020	963220	-	0.003	0.678	0.681	620

negsRNA236	1209520	1209720	-	0.033	0.646	0.679	621
possRNA196	1181321	1181521	+	0.049	0.628	0.677	622
negsRNA190	946695	946895	-	0.039	0.637	0.676	623
negsRNA287	1562260	1562460	-	0.043	0.632	0.675	624
possRNA283	1660921	1661121	+	0.016	0.659	0.675	625
possRNA219	1308624	1308824	+	0.021	0.653	0.675	626
negsRNA198	966750	966950	-	0.102	0.572	0.675	627
possRNA259	1541420	1541620	+	0.026	0.648	0.674	628
negsRNA106	520899	521099	-	0.126	0.548	0.674	629
negsRNA257	1410468	1410668	-	0.029	0.645	0.674	630
possRNA151	952182	952382	+	0.005	0.670	0.674	631
negsRNA253	1384181	1384381	-	0.008	0.665	0.673	632
possRNA216	1300861	1301061	+	0.080	0.592	0.673	633
negsRNA115	539757	539957	-	0.134	0.537	0.671	634
possRNA56	370128	370328	+	0.014	0.654	0.668	635
negsRNA259	1417457	1417657	-	0.052	0.616	0.668	636
negsRNA138	651668	651868	-	0.029	0.636	0.665	637
possRNA80	543320	543520	+	0.021	0.642	0.663	638
possRNA192	1172804	1173004	+	0.021	0.642	0.663	639
possRNA226	1367431	1367631	+	0.022	0.641	0.662	640
negsRNA305	1623938	1624138	-	0.024	0.637	0.661	641
negsRNA130	628835	629035	-	0.104	0.556	0.660	642
negsRNA313	1666854	1667054	-	0.059	0.600	0.660	643
possRNA23	155252	155452	+	0.066	0.592	0.659	644
negsRNA54	258733	258933	-	0.016	0.642	0.658	645
negsRNA348	1843808	1844008	-	0.033	0.625	0.658	646
negsRNA349	1843930	1844130	-	0.018	0.639	0.658	647
negsRNA13	148327	148527	-	0.015	0.642	0.656	648
possRNA127	746130	746330	+	0.062	0.594	0.656	649
possRNA262	1564869	1565069	+	0.025	0.631	0.655	650
possRNA318	1798016	1798216	+	0.122	0.531	0.653	651
possRNA199	1209484	1209684	+	0.059	0.594	0.653	652
negsRNA99	514497	514697	-	0.010	0.642	0.652	653
negsRNA281	1540590	1540790	-	0.000	0.651	0.651	654
possRNA110	653185	653385	+	0.072	0.578	0.650	655

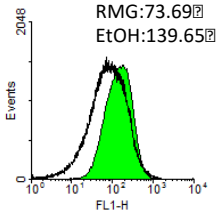
negsRNA155	719727	719927	-	0.003	0.646	0.649	656
negsRNA24	172042	172242	-	0.025	0.621	0.645	657
negsRNA269	1457288	1457488	-	0.099	0.544	0.643	658
negsRNA326	1760711	1760911	-	0.087	0.556	0.643	659
negsRNA111	525889	526089	-	0.104	0.538	0.642	660
negsRNA164	729096	729296	-	0.050	0.589	0.639	661
possRNA228	1367782	1367982	+	0.071	0.563	0.634	662
possRNA307	1781308	1781508	+	0.009	0.622	0.631	663
possRNA154	962039	962239	+	0.017	0.609	0.626	664
possRNA252	1511601	1511801	+	0.052	0.565	0.617	665
possRNA172	1019138	1019338	+	0.035	0.576	0.610	666
negsRNA137	651370	651570	-	0.000	0.607	0.607	667
negsRNA85	412818	413018	-	0.026	0.566	0.593	668
negsRNA56	282499	282699	-	0.019	0.566	0.585	669
possRNA153	960755	960955	+	0.027	0.551	0.578	670
negsRNA274	1511601	1511801	-	0.009	0.565	0.574	671
possRNA128	746277	746477	+	0.057	0.516	0.573	672
possRNA63	412816	413016	+	0.037	0.532	0.570	673
possRNA293	1745289	1745489	+	0.013	0.556	0.569	674
negsRNA331	1794676	1794876	-	0.006	0.556	0.562	675
possRNA159	977643	977843	+	0.013	0.546	0.558	676
negsRNA294	1593619	1593819	-	0.034	0.509	0.543	677
possRNA268	1612707	1612907	+	0.014	0.435	0.449	678

APPENDIX B: SUPPLEMENTARY FIGURES

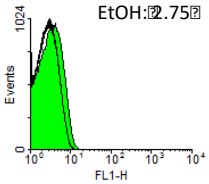
Figure B.1: Flow cytometry histograms for screening all UTR-GFP construct responses to stress.



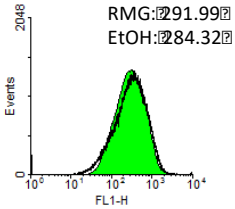
RSE#12



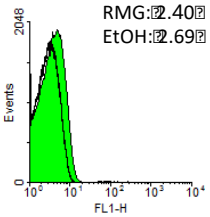
RSE#18



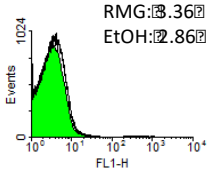
RSE#19



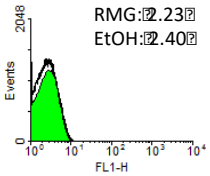
RSE#20



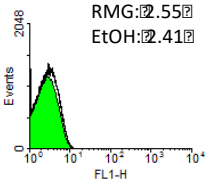
RSE#21



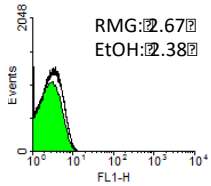
RSE#22



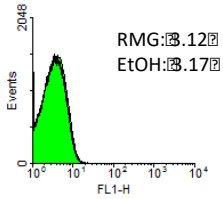
RSE#24



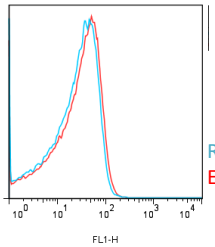
RSE#28



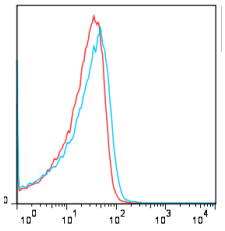
RSE#29



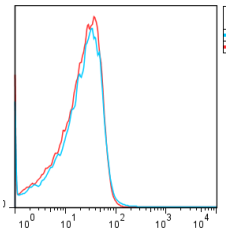
RSE#17



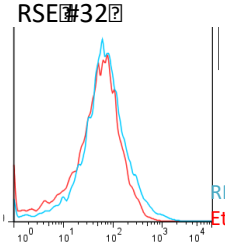
RSE#20



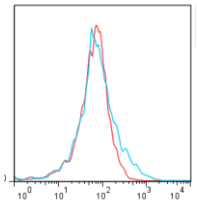
RSE#23



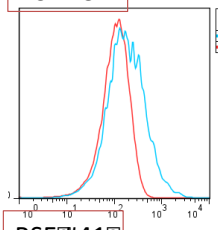
RSE#32



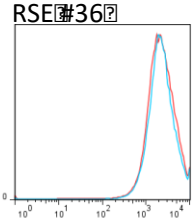
RSE#33



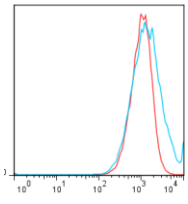
RSE#34



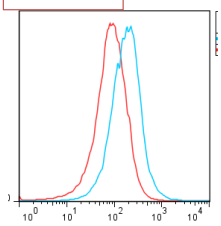
RSE#36



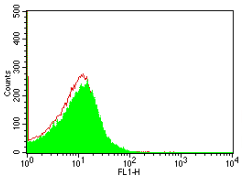
RSE#37



RSE#41

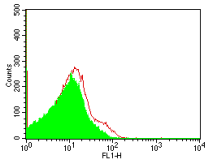


RSE#18



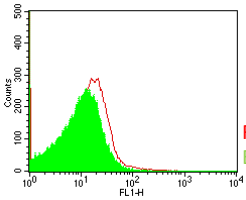
RMG:7.12
ETOH:8.81

RSE#35



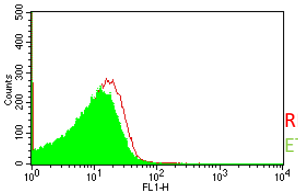
RMG:10.46
ETOH:7.74

RSE#42



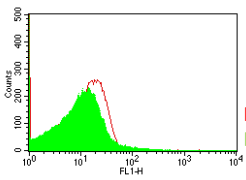
RMG:12.57
ETOH:8.64

RSE#43



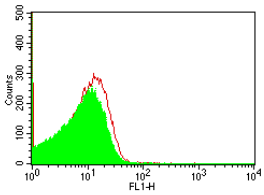
RMG:9.92
ETOH:8.04

RSE#45



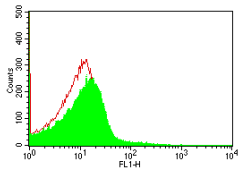
RMG:10.13
ETOH:8.52

RSE#46



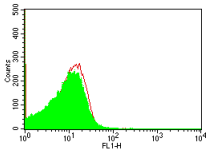
RMG:8.52
ETOH:6.82

RSE#47



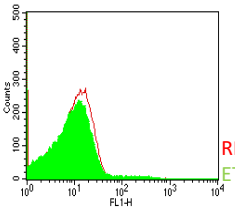
RMG:7.81
ETOH:10.71

RSE#48



RMG:8.64
ETOH:7.75

RSE#50



RMG:8.84
ETOH:7.94

APPENDIX C: SOFTWARE DEVELOPED IN THIS WORK

REFINE (Chapter 4)

User Guide

Starting Point: Transcriptome .bam files for two different conditions (condition1.bam and condition2.bam)

Additional Necessary Files: A reference fasta file for genome (genome.fa) and your initial .gff file (genome.gff)

Necessary Programs: samtools, bedtools

```
awk 'BEGIN{OFS="\t";} $7=="+"' genome.gff > positive_genome.gff

awk 'BEGIN{OFS="\t";} $7=="-"' genome.gff > negative_genome.gff
```

Step 1: Create stranded intergenic .gff files

- Generate an index .fai file from your reference fasta file using the following command:

  ```
  samtools faidx genome.fa
  ```

- Take your new file (genome.fa.fai) and use it to create a genome .txt file

  ```
  awk -v OFS='\t' {'print $1,$2'} genome.fa.fai > genomeFile.txt
  ```

- Separate your .gff file by direction into two separate files

  ```
  awk 'BEGIN{OFS="\t";} $7=="+"' genome.gff > positive_genome.gff

  awk 'BEGIN{OFS="\t";} $7=="-"' genome.gff > negative_genome.gff
  ```

- Use bedtools complement to find the intergenic regions

  ```
  bedtools complement -i positive_genome.gff -g genomeFile.txt > pos_genome_igr.gff

  bedtools complement -i negative_genome.gff -g genomeFile.txt > neg_genome_igr.gff
  ```

Step 2: Convert your .bam files to .txt tab-delimited file format

- Take your two .bam files and run the following:

```
bedtools genomecov -strand + -ibam condition1.bam -d > pos_output1.txt
bedtools genomecov -strand - -ibam condition1.bam -d > neg_output1.txt
bedtools genomecov -strand + -ibam condition2.bam -d > pos_output2.txt
bedtools genomecov -strand - -ibam condition2.bam -d > neg_output2.txt
```

Step 3: Use sRNAScout to locate predicted regions

- Open files.R. Edit the file so that each line contains the ABSOLUTE path to all necessary files.
- Pick appropriate variables for your dataset. You can pick your own variables or run suggest_vars() to get recommended values.
    - min_exp = Minimum level of expression required in a region
    - length_min_exp = Number of nucleotides in a row that a region must be above the minimum expression level.
    - min_de = Minimum differential expression level between reads
    - length_de = Number of reads in a row that must be above the minimum differential expression
    - OPTIONAL: reads_length1 = TOTAL number of reads in your initial condition 1 file. Used for normalization
    - OPTIONAL: reads_length2 = TOTAL number of reads in your initial condition 1 file. Used for normalization.
    - OPTIONAL: distance = how near two regions need to be together to be combined into one continuous region.
- Create the intaRNA input query file using the finalize() function with the variables you have chosen as well as the following file paths:
    - fastaFile = /path/to/genome.fa
    - OPTIONAL: /path/for/output/file. Default is ~/intaRNAInputs.fasta.

- Create the intaRNA target file using the intarna_fasta() function with the following inputs:
    - fastaFile = /path/to/genome.fa
    - gene_names = /path/to/gene_names.txt
    - gene_locs = /path/to/gene_start_locs.txt
    - OPTIONAL: /path/for/output/file

Step 4: Run IntaRNA

- If you are running a large number of regions, generate the accessibility data:

```
mkdir accessibility_dir

IntaRNA -q GAAAACCGCTGAAGATTACCTGGGTGAACCGGTAACTGAAGCTGTTATCAC-t
/path/to/target.fasta --out=tPu:accessibility_dir/intarna.target.pu
```

- Run IntaRNA using the following command:

```
/path/to/IntaRNA -t /path/to/target/fasta -q /path/to/query/fasta --outmode=C --
outCsvCols='id1,start1,end1,id2,start2,end2,E' >> IntaRNAOut.csv
```

Step 5: Calculate the sRNAScores and transScores

- Create a custom gff using the finalize() function in func2_dplyr.R
- Generate htcount count files using two name-sorted sam files and your custom gff:

```
htseq-count -m intersection-nonempty -s reverse -i gene_id sorted_sam_file_1.sam custom_gff.gff >
htcount_1.txt

htseq-count -m intersection-nonempty -s reverse -i gene_id sorted_sam_file_2.sam custom_gff.gff >
htcount_2.txt
```

- Run the 2016-08-31 DESeq2_aero_anaero.R script with the two new htcount files to generate an Excel sheet with information for the transScores.
- Run IntaCalc.R using the output from the 2016-08-31 DESeq2_aero_anaero.R and the output from intaRNA.

- Sort the outputs from IntaCalc to get the top 50 sRNAScores or transScores.

files.R

```
# The location of this folder.
workingDir = "D:/REFINE"

setwd(workingDir)

#The name of your genome to be used for descriptive file names.
template="k12"

fastaFile="ZM4.fa"

gffFile="D:/k12_srna.gff"

# Positive output from bedtools genomecov for condition 1
pos_out_1="D:/sRNApredict_MEG/Intergenic_Regions/pos_k12_output1.txt"

# Positive output from bedtools genomecov for condition 2
pos_out_2="D:/sRNApredict_MEG/Intergenic_Regions/pos_k12_output2.txt"

# Negative output from bedtools genomecov for condition 1
neg_out_1="D:/sRNApredict_MEG/Intergenic_Regions/neg_k12_output1.txt"

# Negative output from bedtools genomecov for condition 2
neg_out_2="D:/sRNApredict_MEG/Intergenic_Regions/neg_k12_output2.txt"

# Positive intergenic gff from bedtools complement
pos_inter_gff="D:/sRNApredict_MEG/Intergenic_Regions/k12_pos_intergenic.gff"

# Negative intergenic gff from bedtools complement
neg_inter_gff="D:/sRNApredict_MEG/Intergenic_Regions/k12_neg_intergenic.gff"

source("func2_dplyr.R")
source("toolBox.R")
```

func2_dplyr.R

```
library("dplyr")
library(magrittr)
library(GenomicRanges)
```

```
source("https://bioconductor.org/biocLite.R")
biocLite("Biostrings")
require(Biostrings)
source("D:/sRNApredict_MEG/Intergenic_Regions/OutputToInter.R")
source("D:/sRNApredict_MEG/sRNA_Predict2/toolBox.R")

pos_inter <- intergen(pos_out_1, pos_out_2, pos_inter_gff)
neg_inter <- intergen(neg_out_1, neg_out_2, neg_inter_gff)

#pos_inter1 <- read.csv("D:/sRNApredict_MEG/sRNA_Predict2/pos_k12_inter1.csv",
header=TRUE, sep = ",", na.strings=c("","NA"), stringsAsFactors=FALSE, colClasses =
c("character", "character", "integer", "integer"))
#pos_inter2 <- read.csv("D:/sRNApredict_MEG/sRNA_Predict2/pos_k12_inter2.csv",
header=TRUE, sep = ",", na.strings=c("","NA"), stringsAsFactors=FALSE, colClasses =
c("character", "character", "integer", "integer"))
#neg_inter1 <- read.csv("D:/sRNApredict_MEG/sRNA_Predict2/neg_k12_inter1.csv",
header=TRUE, sep = ",", na.strings=c("","NA"), stringsAsFactors=FALSE, colClasses =
c("character", "character", "integer", "integer"))
#neg_inter2 <- read.csv("D:/sRNApredict_MEG/sRNA_Predict2/neg_k12_inter2.csv",
header=TRUE, sep = ",", na.strings=c("","NA"), stringsAsFactors=FALSE, colClasses =
c("character", "character", "integer", "integer"))
directions <- c("+", "-")

#function with arguments
# a - number for designated high expression
# b - length of high expression area to qualify region
# c - number fold to look for between aerobic and anaerobic differential expression
# d - length of differential expression to qualify region
# returns a list of the row numbers
higherThan <- function(min_exp, length_min_exp, min_de, length_de, direction,
reads_length1=NA, reads_length2=NA, distance=50, utr_region_length=NA,
max_length=500){

  if (max_length < length_min_exp){
    warning("The entered value for length_min_exp is greater than the set maximum of ",
max_length, ". Maximum changed to ", 2*length_min_exp, ".")
    max_length=length_min_exp*2
  }

  # clean data
  if (direction == "+") {
    inter1 <- filter(pos_inter$inter1, !grepl(".", V2, fixed = TRUE))   # removes duplicate
rows generated by OutputToInter.R. These are labeled as nt 1.1, 2.1, 3.1, etc
```

```
  inter2 <- filter(pos_inter$inter2, !grepl(".", V2, fixed = TRUE))
  inter1 <- inter1[which(!is.na(inter1$V2)),]
  inter2 <- inter2[which(!is.na(inter2$V2)),]
 } else {
  inter1 <- filter(neg_inter$inter1, !grepl(".", V2, fixed = TRUE))   # removes duplicate
rows generated by OutputToInter.R. These are labeled as nt 1.1, 2.1, 3.1, etc
  inter2 <- filter(neg_inter$inter2, !grepl(".", V2, fixed = TRUE))
  inter1 <- inter1[which(!is.na(inter1$V2)),]
  inter2 <- inter2[which(!is.na(inter2$V2)),]
 }

 if (!is.na(reads_length1) && !is.na(reads_length2)){
  norm_ratio = reads_length1/reads_length2
  inter2$V3 <- inter2$V3 * norm_ratio
 } else {
  norm_ratio = 1
 }

 inter <- left_join(inter1, inter2, by = c("V2", "V1", "V2")) %>%
  na.omit() %>%
  dplyr::rename(nt = V2) %>%              # nucleotide number
  dplyr::rename(chr = V1) %>%             # chromosome name
  dplyr::rename(Exp1 = V3.x) %>%          # Expression level, sample 1
  dplyr::rename(Exp2 = V3.y) %>%          # Expression level, sample 1
  select(nt, chr, Exp1, Exp2)   # get rid of duplicate columns by selecting the renamed
ones
 inter$nt <- as.numeric(inter$nt)    # convert nt from chr to numeric to allow sorting


 # differential expression filters
 #differential fold minimum
 de1 <- filter(inter, min_de*inter$Exp1 <= inter$Exp2 & inter$Exp2 != 0)
 de2 <- filter(inter, inter$Exp1/min_de >= inter$Exp2 & inter$Exp1 != 0)
 de <- bind_rows(de1, de2) %>%
  arrange(nt)

 # minimum expression level filter
 min_exp_filt <- filter(inter, inter$Exp1 >= min_exp | inter$Exp2 >= min_exp)

 # first nt of each sequential region
 region_starts_de <- c(1, which(diff(de$nt, lag = 1) != 1), length(de$nt))
 inter_starts <- inter$nt[c(1, which(diff(inter$nt, lag=1)!=1), length(inter$nt))+1]
```

```
  # find regions that may contain UTR's
  utr_regions <- c()
  if (!is.na(utr_region_length)){
    inter_ends <- inter$nt[c(which(diff(inter$nt, lag=1)!=1), length(inter$nt))]
    for (i in 1:length(inter_starts)){
      if(is.na(inter_starts[[i]])){
        print("True")
        break
      }
      utr_regions <- c(utr_regions,
                       seq(inter_starts[[i]], inter_starts[[i]]+utr_region_length),
                       seq(inter_ends[[i]]-utr_region_length, inter_ends[[i]]))
    }
  }

  # sequential diff exp nt in list of lists
  de_regions <- sapply(seq(length(region_starts_de) - 1),  function(i)
de$nt[(region_starts_de[i] + 1):region_starts_de[i+1]])

  # gather only regions >= length_de
  de_regions_of_length = list()
  for (x in de_regions){
    if(length(x) >= length_de){
      de_regions_of_length = c(de_regions_of_length, list(x))
    }
  }

  # find sequential regions with min exp for at least length_min_exp
  region_starts_min <- c(1, which(diff(min_exp_filt$nt, lag = 1) != 1),
length(min_exp_filt$nt))

  # sequential min exp nt in list of lists
  min_exp_regions <- sapply(seq(length(region_starts_min) - 1),  function(i)
min_exp_filt$nt[(region_starts_min[i] + 1):region_starts_min[i+1]])

  # gather only regions >= length_min_exp and splits up large regions at their lowest
trough
  min_exp_regions_of_length = list()
  for (x in min_exp_regions){
    last_min_loc = 1
    if(length(x) >= length_min_exp && length(x) < max_length){
       min_exp_regions_of_length = c(min_exp_regions_of_length, list(x))
    } else if (length(x) > max_length) {
```

```
    last_min = x[1]
    chunk_size = as.integer((2*length(x)/max_length))
    # finds lowest point in each split up region
    for (step in 0:chunk_size){
     start_loc = step*max_length/2 + 1
     end_loc = min(((step+1)*max_length/2 + 1), length(x))
     nt_locs = which(inter$nt %in% x[start_loc:end_loc])
     temp_nt = inter$nt[nt_locs]
     min_loc = which.min(inter$Exp1[nt_locs])
     if (temp_nt[min_loc] > last_min) {
      min_exp_regions_of_length = c(min_exp_regions_of_length, list(seq(last_min,
temp_nt[min_loc])))
      last_min = temp_nt[min_loc] + 1
     }
    }
  }
 }

 # make GRanges objects

 de_granges <- data.frame(chr = character(length(de_regions_of_length)),
                          start = integer(length(de_regions_of_length)),
                          end = integer(length(de_regions_of_length)),
                          strand = character(length(de_regions_of_length)),
                          stringsAsFactors = FALSE)

 for (i in 1:length(de_regions_of_length)) {
  de_granges$chr[i] <- "chromosome"
  de_granges$strand[i] <- "*"
  de_granges$start[i] <- head(de_regions_of_length[[i]], 1)
  de_granges$end[i] <- tail(de_regions_of_length[[i]], 1)
 }

 min_exp_granges <- data.frame(chr = character(length(min_exp_regions_of_length)),
                               start = integer(length(min_exp_regions_of_length)),
                               end = integer(length(min_exp_regions_of_length)),
                               strand = character(length(min_exp_regions_of_length)),
                               stringsAsFactors = FALSE)

 for (i in 1:length(min_exp_regions_of_length)) {
  min_exp_granges$chr[i] <- "chromosome"
  min_exp_granges$strand[i] <- "*"
  min_exp_granges$start[i] <- head(min_exp_regions_of_length[[i]], 1)
```

```
  min_exp_granges$end[i] <- tail(min_exp_regions_of_length[[i]], 1)
 }

 de_GR <- makeGRangesFromDataFrame(de_granges)
 min_exp_GR <- makeGRangesFromDataFrame(min_exp_granges)
 distn = distanceToNearest(min_exp_GR, de_GR)
 hits <- as.data.frame(distn)

 hits_filt <- hits %>%
  filter(distance <= 0)

 sRNA_pred_regions <- data.frame(chr = character(nrow(hits_filt) + 1),
                                 start = integer(nrow(hits_filt) + 1),
                                 end = integer(nrow(hits_filt) + 1),
                                 mid = integer(nrow(hits_filt) + 1),
                                 strand = character(nrow(hits_filt) + 1),
                                 stringsAsFactors = FALSE)

 final_sRNA_pred_regions <- list()
 i_count=1
 for (i in 1:length(hits_filt$queryHits)) {
  region_start = min(min_exp_granges$start[hits_filt$queryHits[i]],
de_granges$start[hits_filt$subjectHits[i]])
  region_end = min(min_exp_granges$end[hits_filt$queryHits[i]],
de_granges$end[hits_filt$subjectHits[i]])
  if (!region_start%in%utr_regions && !region_end%in%utr_regions){
   final_sRNA_pred_regions$chr[i_count] <- "chromosome"
   final_sRNA_pred_regions$strand[i_count] <- "*"
   final_sRNA_pred_regions$start[i_count] <- region_start
   final_sRNA_pred_regions$end[i_count] <- region_end
   i_count=i_count+1
  }
 }

 final_sRNA_pred_regions$start[final_sRNA_pred_regions$start < 0] <- 0

 bed_sRNA <- vector()

 # Combines nearby regions with reasonably similar expression levels
 i = 1
 while (i <= length(final_sRNA_pred_regions$start)){
  final_end = final_sRNA_pred_regions$end[length(final_sRNA_pred_regions$end)]
  i_count = 0
```

```
   curr_start_loc = which(inter1$V2 %in% final_sRNA_pred_regions$start[i])
   curr_end_loc = which(inter1$V2 %in% final_sRNA_pred_regions$end[i])
   new_max = max(inter1$V3[curr_start_loc:curr_end_loc])
   new_min = new_max
   while (i+i_count < length(final_sRNA_pred_regions$start) &&
(final_sRNA_pred_regions$start[i+i_count+1] -
final_sRNA_pred_regions$end[i+i_count]) < distance){
     inter_end_loc = which(inter1$V2 %in%
final_sRNA_pred_regions$start[i+i_count+1])
     peak_end_loc = which(inter1$V2 %in%
final_sRNA_pred_regions$end[i+i_count+1])
     new_max = max(new_max, max(inter1$V3[inter_end_loc:peak_end_loc]))
     new_min = min(new_min, max(inter1$V3[inter_end_loc:peak_end_loc]))
     if (new_min/new_max <= 0.02){
       break
     }
     i_count = i_count + 1
   }

   bed_sRNA <- c(bed_sRNA, seq.int(final_sRNA_pred_regions$start[i],
final_sRNA_pred_regions$end[i+i_count]))

     i = i + i_count + 1
   }

 outputlist = which(inter1$V2 %in% bed_sRNA)
 output <- list("outputlist"=outputlist, "inter1_var"=inter1, "inter2_var"=inter2)
 temp_output <- inter1$V2[outputlist]
 return(output)
}


#function that takes argument of row numbers, alters inter1 and returns a new table with
just those rows
tableOne <- function(input_list){
 inter1_var = input_list$inter1
 x = input_list$outputlist
 newinter1 = inter1_var[x, ]
 newinter1 = newinter1[,3:4]
 colnames(newinter1) = c("loc", "expr")
 return (newinter1)
}
```

```
#function that takes argument of row numbers, alters inter2, and returns a new table with
just those rows
tableTwo <- function(input_list){
  inter2_var = input_list$inter2_var
  x = input_list$outputlist
  newinter2 = inter2_var[x,]
  newinter2 = newinter2[,3:4]
  colnames(newinter2) = c("loc", "expr")
  return (newinter2)
}

#takes in a data frame with "loc" heading
#returns a list of lists containing the sequential regions
#used in num() min() and max() functions in Server.R
getRange <- function(table){
  interlist1 = table$loc
  interBreaks1a <- c(0, which(diff(interlist1, lag = 1) != 1), length(interlist1))
  interBreaks1b = sapply(seq(length(interBreaks1a) - 1),  function(y)
interlist1[(interBreaks1a[y] + 1):interBreaks1a[y+1]])
  return(interBreaks1a)
}
```

toolbox.R

```
htcount <- function(min_exp, length_min_exp, min_de, length_de, outFile){
  pos_higher = higherThan(min_exp, length_min_exp, min_de, length_de, "+")
  neg_higher = higherThan(min_exp, length_min_exp, min_de, length_de, "-")
  pos_output = getRange(tableOne(pos_higher))
  neg_output = getRange(tableOne(neg_higher))

  for (i in 1:length(neg_output)){
    output_locs = which(neg_higher$inter1_var$V2 %in% neg_output[[i]])
    print(sum(neg_higher$inter1_var$V3[output_locs]))
    center = as.integer(length(neg_output[[i]])/2)
    dnastr=""
    if (length(neg_output[[i]]) > 200){
      for (j in 1:length(neg_output[[i]])){

        dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][j]],sep="")
      }
```

```
    i_count = i_count + 1
   } else {
    if (length(neg_output[[i]]) > 6){
     for (j in -100:100){
      dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][center]+j],sep="")
     }
     i_count = i_count + 1
    }
   }
  }

  i_count = 1
  for (i in 1:length(pos_output)){
   dnastr=""
   center = as.integer(length(pos_output[[i]])/2)
   if (length(pos_output[[i]]) > 200){
    for (j in 1:length(pos_output[[i]])){
     dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][j]],sep="")
    }
    i_count = i_count + 1
   } else {

    for (j in -100:100){
     dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][center]+j],sep="")
    }
    i_count = i_count+1
   }
   if (any(pos_output[[i]] %in% flattened_pos_list)){
    lapply()
   }

   lapply(dnastr, write, outFile)
  }
}

# TODO: add a descriptor variable to add a description to the file name
finalize <- function(min_exp, length_min_exp, min_de, length_de, fastaFile) {
 neg_output_full <- list()
 myFastaFile = readDNAStringSet(fastaFile)
 sequence = paste(myFastaFile)
 sequence = strsplit(sequence, NULL)
 pos_output = getRange(tableOne(higherThan(min_exp, length_min_exp, min_de,
length_de, "+")))
```

```
 neg_output = getRange(tableOne(higherThan(min_exp, length_min_exp, min_de,
length_de, "-")))
 i_count = 1

 for (i in 1:length(neg_output)){
  if (any(neg_output[[i]] %in% flattened_neg_list)){
   print(i_count)
  }
  center = as.integer(length(neg_output[[i]])/2)
  dnastr=""
  if (length(neg_output[[i]]) > 200){
   for (j in 1:length(neg_output[[i]])){
    dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][j]],sep="")
   }
   # lapply(paste(">NegSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNANegRef.fasta", append=TRUE)
   # lapply(seq(max(0,neg_output[[i]][[center]]-100),neg_output[[i]][[center]]+100,
sep="\n"), write, "D:/EthanolIntaRNAInputs/intaRNANegRef.fasta", append=TRUE)
   # lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNANegRef.fasta",
append=TRUE)
   # lapply(paste(">NegSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNANeg.fasta", append=TRUE)
   # lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNANeg.fasta",
append=TRUE)
   i_count = i_count + 1
  } else {
   if (length(neg_output[[i]]) > 6){
    for (j in -100:100){
     dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][center]+j],sep="")
    }
    # lapply(paste(">NegSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNANegRef.fasta", append=TRUE)
    # lapply(seq(max(0,neg_output[[i]][[center]]-100),neg_output[[i]][[center]]+100,
sep="\n"), write, "D:/EthanolIntaRNAInputs/intaRNANegRef.fasta", append=TRUE)
    # lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNANegRef.fasta",
append=TRUE)
    # lapply(paste(">NegSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNANeg.fasta", append=TRUE)
    # lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNANeg.fasta",
append=TRUE)
    i_count = i_count + 1
   }
   #}
```

```
  }
 }

 i_count = 1
 for (i in 1:length(pos_output)){
  dnastr=""
  center = as.integer(length(pos_output[[i]])/2)
  if (length(pos_output[[i]]) > 200){
   for (j in 1:length(pos_output[[i]])){
    dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][j]],sep="")
   }
   lapply(paste(">PosSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNAPos.fasta", append=TRUE)
   lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNAPos.fasta", append=TRUE)
   lapply(paste(">PosSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta", append=TRUE)
   lapply(seq(max(0,pos_output[[i]][[center]]-100),pos_output[[i]][[center]]+100,
sep="\n"), write, "D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta", append=TRUE)
   lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta",
append=TRUE)
   i_count = i_count + 1
  } else {
   for (j in -100:100){
    dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][center]+j],sep="")
   }
   lapply(paste(">PosSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNAPos.fasta", append=TRUE)
   lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNAPos.fasta", append=TRUE)
   lapply(paste(">PosSeq", i_count, sep=""), write,
"D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta", append=TRUE)
   lapply(seq(max(0,pos_output[[i]][[center]]-100),pos_output[[i]][[center]]+100),
write, "D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta", append=TRUE)
   lapply(dnastr, write, "D:/EthanolIntaRNAInputs/intaRNAPosRef.fasta",
append=TRUE)
   i_count = i_count + 1
  }
  #}
 }
}

finalize2 <- function(min_exp, length_min_exp, min_de, length_de, fastaFile, outFile,
min_size=200) {
```

```
  neg_output_full <- list()
  myFastaFile = readDNAStringSet(fastaFile)
  sequence = paste(myFastaFile)
  sequence = strsplit(sequence, NULL)
  pos_output = getRange(tableOne(higherThan(min_exp, length_min_exp, min_de,
length_de, "+")))
  neg_output = getRange(tableOne(higherThan(min_exp, length_min_exp, min_de,
length_de, "-")))
  i_count = 1

  for (i in 1:length(neg_output)){
    center = as.integer(length(neg_output[[i]])/2)
    dnastr=""
    if (length(neg_output[[i]]) > min_size){
      lapply(paste("Chromosome", "ena", "sRNA", neg_output[[i]][[1]],
neg_output[[i]][[length(neg_output[[i]])]],".","-",".",
                 paste("ID=gene:negsRNA", i, ";description=small_rna;gene_id=negsRNA", i,
sep=""), sep="\t"),
             write, outFile, append=TRUE)
      for (j in 1:length(neg_output[[i]])){
        dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][j]],sep="")
      }
      i_count = i_count + 1

    } else {
      # cuts off regions smaller than 6 nucleotides long because intaRNA does not accept
regions this short.
      if (length(neg_output[[i]]) > 6){
        lapply(paste("Chromosome", "ena", "sRNA", max(0,neg_output[[i]][[center]]-
100),neg_output[[i]][[center]]+100,".","-",".",
                   paste("ID=gene:negsRNA", i, ";description=small_rna;gene_id=negsRNA",
i, sep=""), sep="\t"),
               write, outFile, append=TRUE)
        for (j in -100:100){
          dnastr = paste(dnastr, sequence[[1]][neg_output[[i]][center]+j],sep="")
        }
        i_count = i_count + 1
      }
    }
  }

  i_count = 1
  for (i in 1:length(pos_output)){
```

```
    dnastr=""
    center = as.integer(length(pos_output[[i]])/2)
    if (length(pos_output[[i]]) > min_size){
      lapply(paste("Chromosome", "ena", "sRNA", pos_output[[i]][[1]],
pos_output[[i]][[length(pos_output[[i]])]],".","+",".",
                 paste("ID=gene:possRNA", i, ";description=small_rna;gene_id=possRNA", i,
sep=""), sep="\t"),
             write, "D:/k12_srna.gff", append=TRUE)
      for (j in 1:length(pos_output[[i]])){
        dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][j]],sep="")
      }
      i_count = i_count + 1
    } else {
      lapply(paste("Chromosome", "ena", "sRNA", max(0,pos_output[[i]][[center]]-
100),pos_output[[i]][[center]]+100,".","+",".",
                 paste("ID=gene:possRNA", i,
";biotype=protein_coding;description=small_rna;gene_id=possRNA", i, sep=""),
sep="\t"),
             write, "D:/k12_srna.gff", append=TRUE)
      for (j in -100:100){
        dnastr = paste(dnastr, sequence[[1]][pos_output[[i]][center]+j],sep="")
      }
      i_count = i_count + 1
    }
  }
}


get_gene_locs <- function(gffFile=gffFile){
  gff = read.csv(gffFile, sep="\t", header=FALSE)
  gff = filter(gff, grepl("gene", V3, fixed=TRUE))
  gff = gff[!duplicated(gff$V4),]
  gene_start_locs = gff$V4
  gene_names = as.character(gff$V9)
  gene_names = unlist(strsplit(gene_names, ";"))
  gene_names = gene_names[str_detect(gene_names, "ID=gene:")]
  gene_names = gsub("ID=gene:","", gene_names)
  output = list(start=gene_start_locs, name=gene_names)
}

intarna_fasta <- function(fastaFile=fastaFile, gffFile=gffFile, template=template) {
  myFastaFile = readDNAStringSet(fastaFile)
  sequence = paste(myFastaFile)
```

```
  sequence = strsplit(sequence, NULL)
  gene_data = get_gene_locs(gffFile)
  start_locs = gene_data$start
  gene_names = gene_data$name
  start_locs = start_locs$V1
  end_locs = start_locs + 100
  start_locs = start_locs - 200
  start_locs[start_locs < 0] <- 0
  end_locs[end_locs > length(myFastaFile[[1]])] <- length(myFastaFile[[1]])
  last_start_loc=-1
  for (i in 1:length(start_locs)){
    if (start_locs[[i]]==last_start_loc){
      next
    }
    dnastr=""
    full_loc = seq(start_locs[[i]], end_locs[[i]])
    for (j in 1:length(full_loc)){
      dnastr = paste(dnastr, sequence[[1]][full_loc[[j]]],sep="")
    }
    lapply(paste(">",gene_names$V1[i],sep=""), write, paste(template, "fasta.fasta",
sep=""), append=TRUE)
    lapply(dnastr, write, paste(template, "fasta.fasta", sep=""), append=TRUE)
    last_start_loc=start_locs[[i]]
  }
}
```

run_DESeq2.R

```
library("DESeq2")
countdata <- read.csv("C:/Users/kh23462/Dropbox/Contreras
Group/REFINE/counts_aero_anaero.csv", header = TRUE, row.names=1)
id <- colnames(countdata)
strain <- c("aero", "anaero")
coldata <- data.frame(row.names=id,strain) #table of sample descriptions (w ID and
factors)
coldata$strain = factor(x = coldata$strain, levels = c('aero', 'anaero'))
coldata$strain <- relevel(coldata$strain, "aero")
coldata
```

```
ddsMat <- DESeqDataSetFromMatrix(countData = countdata, colData = coldata, design
= ~ strain)


#Run DESeq
ddsTC <- DESeq(ddsMat, test="Wald")
resTC <- results(ddsTC)
resTC$symbol <- mcols(ddsTC)$symbol
head(resTC[order(resTC$padj),],4)
data <- plotCounts(ddsTC, which.min(resTC$padj),
                   intgroup=c("strain"), returnData=TRUE)

#Exporting results
resOrdered <- resTC[order(resTC$padj),]
head(resOrdered)
resOrderedDF <- as.data.frame(resOrdered)
library("xlsx")
write.xlsx2(resOrderedDF, file="2016-08-31 DESeq2_aero_anaero.xlsx")

#exporting specific log2fold changes
res2 <- results(ddsTC, contrast=c("strain","aero","anaero"))
write.xlsx2(res2, file="2016-08-31 DESeq2_aero_anaero.xlsx", sheetName =
"aero_anaero_wald", append = TRUE)
```

References

1. Klinke, H. B., Thomsen, A. B. & Ahring, B. K. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* **66,** 10–26 (2004).
2. Stephanopoulos, G. Challenges in engineering microbes for biofuels production. *Science* **315,** 801–4 (2007).
3. Patnaik, R. Engineering complex phenotypes in industrial strains. *Biotechnol. Prog.* **24,** 38–47
4. Hoe, C.-H., Raabe, C. A., Rozhdestvensky, T. S. & Tang, T.-H. Bacterial sRNAs: Regulation in stress. *Int. J. Med. Microbiol.* **303,** 217–229 (2013).
5. Gottesman, S. *et al.* Small RNA Regulators and the Bacterial Response to Stress. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 1–11 (2006).
6. Bobrovskyy, M. & Vanderpool, C. K. Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms. *Annu. Rev. Genet.* **47,** 209–232 (2013).
7. Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.* **3,** (2007).
8. Hershberg, R., Altuvia, S. & Margalit, H. A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.* **31,** 1813–1820 (2003).
9. Haning, K., Cho, S. H. & Contreras, L. M. Small RNAs in mycobacteria: an unfolding story. *Front. Cell. Infect. Microbiol.* **4,** (2014).
10. Gripenland, J. *et al.* RNAs: regulators of bacterial virulence. *Nat. Rev. Microbiol.* **8,** 857–866 (2010).
11. Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nat. Rev. Microbiol.* **10,** 255–265 (2012).
12. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43,** 880–91 (2011).
13. Romeo, T., Vakulskas, C. A. & Babitzke, P. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems: The Csr global regulatory system. *Environ. Microbiol.* **15,** 313–324 (2013).
14. Lee, H.-J. & Gottesman, S. sRNA roles in regulating transcriptional regulators: Lrp and SoxS regulation by sRNAs. *Nucleic Acids Res.* **44,** 6907–6923 (2016).
15. Zhang, A. *et al.* Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50,** 1111–24 (2003).
16. Na, D. *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31,** 170–174 (2013).
17. Vazquez-Anderson, J. & Contreras, L. M. Regulatory RNAs: Charming gene management styles for synthetic biology applications. *RNA Biol.* **10,** 1778–1797 (2013).
18. Chappell, J. *et al.* The centrality of RNA for engineering gene expression. *Biotechnol. J.* **8,** 1379–95 (2013).
19. Kang, Z. *et al.* Small RNA regulators in bacteria: powerful tools for metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* **98,** 3413–24

(2014).
20. Qi, L. S. & Arkin, A. P. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat. Rev. Microbiol.* **12,** 341–54 (2014).
21. Mellin, J. R. & Cossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31,** 150–156 (2015).
22. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23,** 995–1011 (2017).
23. Villa, J. K., Su, Y., Contreras, L. M. & Hammond, M. C. Synthetic Biology of Small RNAs and Riboswitches. *Microbiol. Spectr.* **6,** (2018).
24. Haning, K., Cho, S. H. & Contreras, L. M. Strain engineering via regulatory noncoding RNAs: not a one-blueprint-fits-all. *Curr. Opin. Chem. Eng.* **10,** 25–34 (2015).
25. Nakashima, N. & Miyazaki, K. Bacterial cellular engineering by genome editing and gene silencing. *Int. J. Mol. Sci.* **15,** 2773–93 (2014).
26. Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription activating RNAs. *Nat. Chem. Biol.* **advance on,** (2015).
27. Gaida, S. M., Al-Hinai, M. a, Indurthi, D. C., Nicolaou, S. a & Papoutsakis, E. T. Synthetic tolerance: three noncoding small RNAs, DsrA, ArcZ and RprA, acting supra-additively against acid stress. *Nucleic Acids Res.* **41,** 8726–37 (2013).
28. Weber, H., Polen, T., Heuveling, J., Wendisch, V. F. & Hengge, R. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J. Bacteriol.* **187,** 1591–603 (2005).
29. Dong, T. & Schellhorn, H. E. Role of RpoS in Virulence of Pathogens. *Infect. Immun.* **78,** 887–897 (2009).
30. Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-mediated general stress response in Escherichia coli. *Annu. Rev. Microbiol.* **65,** 189–213 (2011).
31. Doelle, H. W., Kirk, L., Crittenden, R., Toh, H. & Doelle, M. B. *Zymomonas Mobilis* —Science and Industrial Application. *Crit. Rev. Biotechnol.* **13,** 57–98 (1993).
32. He, M. *et al.* Zymomonas mobilis: a novel platform for future biorefineries. *Biotechnol. Biofuels* **7,** (2014).
33. Panesar, P. S., Marwaha, S. S. & Kennedy, J. F. Zymomonas mobilis: an alternative ethanol producer. *J. Chem. Technol. Biotechnol.* **81,** 623–635 (2006).
34. Rogers, P. L., Jeon, Y. J., Lee, K. J. & Lawford, H. G. Zymomonas mobilis for fuel ethanol and higher value products. *Adv. Biochem. Eng. Biotechnol.* **108,** 263–88 (2007).
35. Yang, S. *et al. Zymomonas mobilis* as a model system for production of biofuels and biochemicals. *Microb. Biotechnol.* (2016). doi:10.1111/1751-7915.12408
36. Swings, J. & De Ley, J. The biology of Zymomonas. *Bacteriol. Rev.* **41,** 1–46 (1977).
37. Bajpai, P. K. & Margaritis, A. Effect of temperature and pH on immobilized Zymomonas mobilis for continuous production of ethanol. *Biotechnol. Bioeng.* **28,**

824–828 (1986).
38. Rogers, P. L., Lee, K. J., Skotnicki, M. L. & Tribe, D. E. in 37–84 (Springer, Berlin, Heidelberg, 1982). doi:10.1007/3540116982_2
39. Yang, Y. *et al.* Progress and perspective on lignocellulosic hydrolysate inhibitor tolerance improvement in Zymomonas mobilis. *Bioresour. Bioprocess.* **5,** 6 (2018).
40. Tan, F. *et al.* Using global transcription machinery engineering (gTME) to improve ethanol tolerance of Zymomonas mobilis. *Microb. Cell Fact.* **15,** 4 (2016).
41. Yang, S. *et al.* Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in Zymomonas mobilis and Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10395–400 (2010).
42. Tao, F., Miao, J. Y., Shi, G. Y. & Zhang, K. C. Ethanol fermentation by an acid-tolerant Zymomonas mobilis under non-sterilized condition. *Process Biochem.* **40,** 183–187 (2005).
43. Shui, Z.-X. *et al.* Adaptive laboratory evolution of ethanologenic Zymomonas mobilis strain tolerant to furfural and acetic acid inhibitors. *Appl. Microbiol. Biotechnol.* **99,** 5739–5748 (2015).
44. Tan, F.-R. *et al.* Improving furfural tolerance of Zymomonas mobilis by rewiring a sigma factor RpoD protein. *Appl. Microbiol. Biotechnol.* (2015). doi:10.1007/s00253-015-6577-2
45. Wang, J.-L. *et al.* Engineered Zymomonas mobilis for salt tolerance using EZ-Tn5-based transposon insertion mutagenesis system. *Microb. Cell Fact.* **15,** 101 (2016).
46. Zhao, N. *et al.* Flocculating *Zymomonas mobilis* is a promising host to be engineered for fuel ethanol production from lignocellulosic biomass. *Biotechnol. J.* **9,** 362–371 (2014).
47. Seo, J.-S. *et al.* The genome sequence of the ethanologenic bacterium Zymomonas mobilis ZM4. *Nat. Biotechnol.* **23,** 63–8 (2005).
48. Yang, S. *et al.* Transcriptomic and metabolomic profiling of Zymomonas mobilis during aerobic and anaerobic fermentations. *BMC Genomics* **10,** 34 (2009).
49. Wang, X. *et al.* Advances and prospects in metabolic engineering of Zymomonas mobilis. *Metab. Eng.* (2018). doi:10.1016/j.ymben.2018.04.001
50. He, M. *et al.* Transcriptome profiling of Zymomonas mobilis under furfural stress. *Appl. Microbiol. Biotechnol.* **95,** 189–99 (2012).
51. Yang, S. *et al.* Elucidation of Zymomonas mobilis physiology and stress responses by quantitative proteomics and transcriptomics. *Front. Microbiol.* **5,** 246 (2014).
52. Yang, S. *et al.* Systems Biology Analysis of Zymomonas mobilis ZM4 Ethanol Stress Responses. *PLoS One* **8,** e68886 (2013).
53. Cho, S. H., Lei, R., Henninger, T. D. & Contreras, L. M. Discovery of Ethanol-Responsive Small RNAs in *Zymomonas mobilis*. *Appl. Environ. Microbiol.* **80,** 4189–4198 (2014).
54. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a

new approach for improving cellular phenotype. *Metab. Eng.* **9,** 258–67 (2007).
55. Chen, Y., Indurthi, D. C., Jones, S. W. & Papoutsakis, E. T. Small RNAs in the genus Clostridium. *MBio* **2,** e00340-10 (2011).
56. Bi, Y. *et al.* Regulation Mechanism Mediated by Trans-Encoded sRNA Nc117 in Short Chain Alcohols Tolerance in Synechocystis sp. PCC 6803. *Front. Microbiol.* **9,** 863 (2018).
57. Wiegand, S. *et al.* RNA-Seq of Bacillus licheniformis: active regulatory RNA features expressed within a productive fermentation. *BMC Genomics* **14,** 667 (2013).
58. Doran-Peterson, J., Cook, D. M. & Brandon, S. K. Microbial conversion of sugars from plant biomass to lactic acid or ethanol. *Plant J.* **54,** 582–592 (2008).
59. Mohagheghi, A. *et al.* Improving xylose utilization by recombinant Zymomonas mobilis strain 8b through adaptation using 2-deoxyglucose. *Biotechnol. Biofuels* **7,** 19 (2014).
60. Zhang, M., Eddy, C., Deanda, K., Finkelstein, M. & Picataggio, S. Metabolic engineering of a pentose metabolism pathway in ethanologenic Zymomonas mobilis. *Science (80-.).* **267,** 240–243 (1995).
61. Zhang, M. *et al.* Zymomonas pentose-sugar fermenting strains and uses thereof. 1–63 (2007).
62. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15,** 423–437 (2014).
63. He, M.-X. *et al.* Transcriptome profiling of Zymomonas mobilis under ethanol stress. *Biotechnol. Biofuels* **5,** 75 (2012).
64. Yang, S. *et al.* Insights into acetate toxicity in Zymomonas mobilis 8b using different substrates. *Biotechnol. Biofuels* **7,** 140 (2014).
65. Yang, S. *et al.* Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in Zymomonas mobilis and Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10395–400 (2010).
66. Beisel, C. L. *et al.* Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.* **34,** 866–882 (2010).
67. Toledo-Arana, A. *et al.* The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459,** 950–956 (2009).
68. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. & Altuvia, S. A pH-responsive riboregulator. *Genes Dev.* **23,** 2650–62 (2009).
69. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5,** 451–463 (2004).
70. Soukup, J. K. & Soukup, G. A. Riboswitches exert genetic control through metabolite-induced conformational change. *Curr. Opin. Struct. Biol.* **14,** 344–349 (2004).
71. Roth, A. *et al.* A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14,** 308–317 (2007).
72. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–6 (2002).

73. Winkler, W. C. & Breaker, R. R. Genetic Control by Metabolite-Binding Riboswitches. *ChemBioChem* **4,** 1024–1032 (2003).
74. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8,** R239 (2007).
75. Ko, J. & Altman, S. OLE RNA, an RNA motif that is highly conserved in several extremophilic bacteria, is a substrate for and can be regulated by RNase P RNA. *Proc. Natl. Acad. Sci.* **104,** 7815–7820 (2007).
76. Wallace, J. G., Zhou, Z. & Breaker, R. R. OLE RNA protects extremophilic bacteria from alcohol toxicity. *Nucleic Acids Res.* **40,** 6898–6907 (2012).
77. Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* **22,** 445–452 (2006).
78. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31,** 439–441 (2003).
79. Nawrocki, E. P. *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43,** D130–D137 (2015).
80. Nahvi, A. *et al.* Genetic Control by a Metabolite Binding mRNA. *Chem. Biol.* **9,** 1043–1049 (2002).
81. Baker, J. L. *et al.* Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335,** 233–5 (2012).
82. Stockbridge, R. B. *et al.* Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15289–94 (2012).
83. Yang, S., Pelletier, D., Lu, T.-Y. & Brown, S. The Zymomonas mobilis regulator hfq contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol.* **10,** 135 (2010).
84. Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F. & Backofen, R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18,** 900–14 (2012).
85. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
86. Douka, E., Christogianni, A., Koukkou, A. I., Afendra, A. S. & Drainas, C. Use of a green fluorescent protein gene as a reporter in Zymomonas mobilis and Halomonas elongata. *FEMS Microbiol. Lett.* **201,** 221–227 (2001).
87. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* **32,** 1610–1614 (2004).
88. Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into Their Function. *Chem. Biol.* **14,** 173–184 (2007).
89. Topp, S. *et al.* Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol.* **76,** 7881–4 (2010).
90. Yang, S. *et al.* Metabolic engineering of Zymomonas mobilis for 2,3-butanediol production from lignocellulosic biomass sugars. *Biotechnol. Biofuels* **9,** 189 (2016).

91. Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E. & Gross, C. A. Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. *J. Bacteriol.* **189,** 1963–73 (2007).
92. Torres-Quesada, O. *et al.* Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in *Sinorhizobium meliloti*. *RNA Biol.* **11,** 563–579 (2014).
93. Vecerek, B., Moll, I. & Bläsi, U. Translational autocontrol of the Escherichia coli hfq RNA chaperone gene. *RNA* **11,** 976–984 (2005).
94. Sobrero, P. & Valverde, C. The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* **38,** 276–299 (2012).
95. Ramos, C. G. *et al.* Regulation of Hfq mRNA and Protein Levels in Escherichia coli and Pseudomonas aeruginosa by the Burkholderia cenocepacia MtvR sRNA. *PLoS One* **9,** e98813 (2014).
96. Sowa, S. W. *et al.* Integrative FourD omics approach profiles the target network of the carbon storage regulatory system. *Nucleic Acids Res.* **45,** gkx048 (2017).
97. Oliva, G., Sahr, T. & Buchrieser, C. Small RNAs, 5′ UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiol. Rev.* **39,** 331–349 (2015).
98. Rabhi, M. *et al.* The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.* **30,** 2805–16 (2011).
99. Kohárýová, M. & Kolárová, M. Oxidative stress and thioredoxin system. *Gen. Physiol. Biophys.* **27,** 71–84 (2008).
100. Liu, J. M. & Camilli, A. A broadening world of bacterial small RNAs. *Curr. Opin. Microbiol.* **13,** 18–23 (2010).
101. Christiansen, J. K., Larsen, M. H., Ingmer, H., Søgaard-Andersen, L. & Kallipolitis, B. H. The RNA-binding protein Hfq of Listeria monocytogenes: role in stress tolerance and virulence. *J. Bacteriol.* **186,** 3355–62 (2004).
102. Pichon, C. & Felden, B. Proteins that interact with bacterial small RNA regulators. *FEMS Microbiol. Rev.* **31,** 614–625 (2007).
103. Tsai, C.-H., Liao, R., Chou, B., Palumbo, M. & Contreras, L. M. Genome-Wide Analyses in Bacteria Show Small-RNA Enrichment for Long and Conserved Intergenic Regions. *J. Bacteriol.* **197,** 40–50 (2015).
104. Friedman, R. C. *et al.* Common and phylogenetically widespread coding for peptides by bacterial small RNAs. *BMC Genomics* **18,** 553 (2017).
105. Gottesman, S. & Storz, G. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb. Perspect. Biol.* **3,** a003798–a003798 (2011).
106. Gelderman, G. & Contreras, L. M. in *Systems Metabolic Engineering: Methods and Protocols* (ed. Alper, H. S.) **985,** 269–95 (Humana Press, 2013).
107. Tsai, C.-H., Liao, R., Chou, B., Palumbo, M. & Contreras, L. M. Genome-wide analysis in bacteria show small RNA enrichment for long and conserved intergenic regions. *J. Bacteriol.* JB.02359-14- (2014). doi:10.1128/JB.02359-14
108. Melamed, S. *et al.* Global Mapping of Small RNA-Target Interactions in Bacteria.

Mol. Cell **63,** 884–897 (2016).
109. Lalaouna, D. & Masse, E. Identification of sRNA interacting with a transcript of interest using MS2-affinity purification coupled with RNA sequencing (MAPS) technology. *Genomics Data* **5,** (2015).
110. Busch, A., Richter, A. S. & Backofen, R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24,** 2849–56 (2008).
111. Vogel, J. *et al.* RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31,** 6435–6443 (2003).
112. Fröhlich, K. S., Haneke, K., Papenfort, K. & Vogel, J. The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res.* **41,** gkw632 (2016).
113. Miyakoshi, M., Chao, Y. & Vogel, J. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* **34,** 1478–1492 (2015).
114. Pei, G., Sun, T., Chen, S., Chen, L. & Zhang, W. Systematic and functional identification of small non-coding RNAs associated with exogenous biofuel stress in cyanobacterium Synechocystis sp. PCC 6803. *Biotechnol. Biofuels* **10,** 57 (2017).
115. Block, K. F., Puerta-Fernandez, E., Wallace, J. G. & Breaker, R. R. Association of OLE RNA with bacterial membranes via an RNA-protein interaction: OLE RNA membrane localization. *Mol. Microbiol.* **79,** 21–34 (2011).
116. Said, N. *et al.* In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.* **37,** e133 (2009).
117. Kaboord, B., Smith, S., Patel, B. & Meier, S. in 135–151 (2015). doi:10.1007/978-1-4939-2550-6_12
118. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32,** 170–173 (2011).
119. Vazquez-Anderson, J. *et al.* Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions. *Nucleic Acids Res.* **9,** 104–109 (2017).
120. Yurimoto, H. *et al.* HxlR, a member of the DUF24 protein family, is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible *hxlAB* operon in *Bacillus subtilis*. *Mol. Microbiol.* **57,** 511–519 (2005).
121. Yang, Y. *et al.* Characterization of the Shewanella oneidensis Fur gene: roles in iron and acid tolerance response. *BMC Genomics* **9,** S11 (2008).
122. Nicolaou, S. A., Fast, A. G., Nakamaru-Ogiso, E. & Papoutsakis, E. T. Overexpression of fetA (ybbL) and fetB (ybbM), Encoding an Iron Exporter, Enhances Resistance to Oxidative Stress in Escherichia coli. *Appl. Environ. Microbiol.* **79,** 7210–9 (2013).
123. Yi, X., Gu, H., Gao, Q., Liu, Z. L. & Bao, J. Transcriptome analysis of Zymomonas mobilis ZM4 reveals mechanisms of tolerance and detoxification of phenolic aldehyde inhibitors from lignocellulose pretreatment. *Biotechnol.*

Biofuels **8,** 153 (2015).
124. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447-52 (2015).
125. Zhang, K. *et al.* Transcriptional analysis of adaptation to high glucose concentrations in Zymomonas mobilis.
126. Brown, S. D. *et al.* Mutant alcohol dehydrogenase leads to improved ethanol tolerance in Clostridium thermocellum. doi:10.1073/pnas.1102444108
127. Gibson, J. B. & Wilks, A. V. The alcohol dehydrogenase polymorphism of Drosophila melanogaster in relation to environmental ethanol, ethanol tolerance and alcohol dehydrogenase activity. *Heredity (Edinb).* **60,** 403–414 (1988).
128. Edenberg, H. J. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res. Health* **30,** 5–13 (2007).
129. Wassarman, K. M. Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109,** 141–144 (2002).
130. Li, F. *et al.* Constitutive expression of RyhB regulates the heme biosynthesis pathway and increases the 5-aminolevulinic acid accumulation in Escherichia coli. *FEMS Microbiol. Lett.* **350,** 209–15 (2014).
131. Kang, Z., Wang, X., Li, Y., Wang, Q. & Qi, Q. Small RNA RyhB as a potential tool used for metabolic engineering in Escherichia coli. *Biotechnol. Lett.* **34,** 527–31 (2012).
132. McKee, A. E. *et al.* Manipulation of the carbon storage regulator system for metabolite remodeling and biofuel production in Escherichia coli. *Microb. Cell Fact.* **11,** 79 (2012).
133. Jones, A. J., Venkataramanan, K. P. & Papoutsakis, T. Overexpression of two stress-responsive, small, non-coding RNAs, 6S and tmRNA, imparts butanol tolerance in *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* **363,** fnw063 (2016).
134. Massé, E., Salvail, H., Desnoyers, G. & Arguin, M. Small RNAs controlling iron metabolism. *Curr. Opin. Microbiol.* **10,** 140–5 (2007).
135. Kaur, P. & Balgir, P. P. in *Advances in Animal Biotechnology and its Applications* 3–18 (Springer Singapore, 2018). doi:10.1007/978-981-10-4702-2_1
136. Livny, J. in *Bacterial Regulatory RNA* 3–14 (Humana Press, 2012). doi:10.1007/978-1-61779-949-5_1
137. Washietl, S., Hofacker, I. L. & Stadler, P. F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2454–9 (2005).
138. Herbig, A. & Nieselt, K. nocoRNAc: Characterization of non-coding RNAs in prokaryotes. *BMC Bioinformatics* **12,** 40 (2011).
139. Livny, J., Fogel, M. A., Davis, B. M. & Waldor, M. K. sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.* **33,** 4096–4105 (2005).
140. Sridhar, J. *et al.* sRNAscanner: A Computational Tool for Intergenic Small RNA Detection in Bacterial Genomes. *PLoS One* **5,** e11970 (2010).
141. Pichon, C. & Felden, B. Intergenic sequence inspector: searching and identifying

bacterial RNAs. *Bioinforma. Appl. NOTE* **19,** 1707–1709 (2003).
142. Barquist, L. & Vogel, J. Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu. Rev. Genet.* **49,** 367–394 (2015).
143. Gish, W. R. WU-BLAST. (2002).
144. Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C. & Collins, J. J. Functional characterization of bacterial sRNAs using a network biology approach. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 15522–7 (2011).
145. Papenfort, K. *et al.* Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol. Microbiol.* **68,** 890–906 (2008).
146. Mars, R. A. T. *et al.* Small Regulatory RNA-Induced Growth Rate Heterogeneity of Bacillus subtilis. *PLoS Genet.* **11,** e1005046 (2015).
147. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13,** 227–32 (2012).
148. Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* **39,** D19-21 (2011).
149. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
150. Tsai, C.-H., Liao, R., Chou, B., Palumbo, M. & Contreras, L. M. Genome-Wide Analyses in Bacteria Show Small-RNA Enrichment for Long and Conserved Intergenic Regions. *J. Bacteriol.* **197,** 40–50 (2015).
151. Wright, P. R. *et al.* CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.* **42,** W119-23 (2014).
152. Pain, A. *et al.* An Assessment of Bacterial Small RNA Target Prediction Programs. *RNA Biol.* (2015). doi:10.1080/15476286.2015.1020269
153. Chareyre, S. & Mandin, P. Bacterial Iron Homeostasis Regulation by sRNAs. *Microbiol. Spectr.* **6,** (2018).
154. Jacques, J.-F. *et al.* RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in Escherichia coli. *Mol. Microbiol.* **62,** 1181–1190 (2006).
155. Bak, G. *et al.* Identification of novel sRNAs involved in biofilm formation, motility and fimbriae formation in Escherichia coli. *Sci. Rep.* **5,** 15287 (2015).
156. Salvail, H. *et al.* A small RNA promotes siderophore production through transcriptional and metabolic remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15223–8 (2010).
157. Seo, S. W. *et al.* Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. *Nat. Commun.* **5,** 4910 (2014).
158. Moreau, R. A., Powell, M. J., Fett, W. F. & Whitaker, B. D. News & Notes: The Effect of Ethanol and Oxygen on the Growth of Zymomonas mobilis and the Levels of Hopanoids and Other Membrane Lipids. *Curr. Microbiol.* **35,** 124–128 (1997).

159. Bringer, S., Finn, R. K. & Sahm, H. Effect of oxygen on the metabolism of Zymomonas mobilis. *Arch. Microbiol.* **139,** 376–381 (1984).
160. Mills, T. Y., Sandoval, N. R. & Gill, R. T. Cellulosic hydrolysate toxicity and tolerance mechanisms in Escherichia coli. *Biotechnol. Biofuels* **2,** 26 (2009).
161. Shao, M., Ma, J. & Wang, S. DeepBound: accurate identification of transcript boundaries via deep convolutional neural fields. *Bioinformatics* **33,** i267–i273 (2017).
162. Cho, S. H. *et al.* Identification and Characterization of 5′ Untranslated Regions (5′UTRs) in Zymomonas mobilis as Regulatory Biological Parts. *Front. Microbiol.* **8,** 2432 (2017).
163. Chao, Y. & Vogel, J. The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13,** 24–33 (2010).
164. Kakoschke, T. K. *et al.* The RNA Chaperone Hfq Is Essential for Virulence and Modulates the Expression of Four Adhesins in Yersinia enterocolitica. *Sci. Rep.* **6,** 29275 (2016).
165. Liu, Y. *et al.* Hfq Is a Global Regulator That Controls the Pathogenicity of Staphylococcus aureus. *PLoS One* **5,** e13069 (2010).
166. Vazquez-Anderson, J. *et al.* High throughput in vivo mapping of RNA accessible interfaces to identify functional sRNA binding sites. *Nat. Commun.* (2018).
167. Kast, P. pKSS — A second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138,** 109–114 (1994).
168. Miyazaki, K. Molecular engineering of a PheS counterselection marker for improved operating efficiency in *Escherichia coli*. *Biotechniques* **58,** 86–8 (2015).
169. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **3,** e3647 (2008).
170. Cho, S. H., Contreras, L. M. & Ju, S. H. Synthetic chimeras with orthogonal ribosomal proteins increase translation yields by recruiting mRNA for translation as measured by profiling active ribosomes. *Biotechnol. Prog.* **32,** 285–293 (2016).
171. Zou, S. *et al.* Enhanced electrotransformation of the ethanologen Zymomonas mobilis ZM4 with plasmids: Electrotransformation of Zymomonas mobilis. *Eng. Life Sci.* **12,** 152–161 (2012).
172. Crook, N. *et al.* In vivo continuous evolution of genes and pathways in yeast. *doi.org* **7,** 13051 (2016).
173. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).
174. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. (2016).
175. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
176. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
177. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

178. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
179. Wickham, H. *ggplot2 : elegant graphics for data analysis*. (Springer, 2009).
180. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).
181. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* **5,** e13984 (2010).
182. Takata, K., Reh, S., Tomida, J., Person, M. D. & Wood, R. D. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat. Commun.* **4,** 2338 (2013).
183. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).